Atomic Energy Control Board
P.O. Box 1046
Ottawa, Canada
K1P 5S9

Commission de contrôle de l'énergie atomique
C.P. 1046
Ottawa, Canada
K1P 5S9

INFO-0139

PROBABILISTIC CALCULATION OF DOSE COMMITMENT FROM URANIUM MILL TAILINGS

by

IEC Beak Consultants Ltd.

A research report prepared for the
Atomic Energy Control Board
Ottawa, Canada

RESEARCH REPORT

October 1983

PROBABILISTIC CALCULATION OF DOSE COMMITMENT FROM URANIUM MILL TAILINGS

A research report prepared for the Atomic Energy Control Board by IEC Beak Consultants Ltd., 6870 Goreway Drive, Mississauga, Ontario, L4V 1P1.

ABSTRACT

The report discusses in a general way considerations of uncertainty in relation to probabilistic modelling. An example of a probabilistic calculation applied to the behaviour of uranium mill tailings is given.

RÉSUMÉ

Le rapport discute, d'une manière générale, des réflexions portant sur l'incertitude relative à la modelisation probabiliste. On donne un exemple d'un calcul probabiliste s'appliquant au comportement des résidus provenant d'usines de concentration d'uranium.

DISCLAIMER

The Atomic Energy Control Board is not responsible for the accuracy of the statements made or opinions expressed in this publication and neither the Board nor the author assumes liability with respect to any damage or loss incurred as a result of the use made of the information contained in this publication.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1.0 INTRODUCTION

1.1 Background
1.2 General Prediction Reliability Considerations
1.3 Characterizing Prediction Reliability in Numerical Terms

2.0 APPROACHES TO THE STUDY OF A MATHEMATICAL MODEL'S SENSITIVITY

2.1 Monte Carlo Methods
2.1.1 Fundamental Preliminaries
2.1.2 Some Sampling Plans
2.1.3 Notes Concerning Inference Procedures
2.1.4 Comparison of Techniques Associated with Monte Carlo Method, and Assessment of Monte Carlo Results

2.2 Analytical Methods
2.3 Summary Comment

3.0 ASSESSING THE EXCEEDANCE FUNCTION FOR DOSE

3.1 Estimation of the Exceedance Function
3.2 Properties of the Empirical Exceedance Functions
3.3 The Exceedance Functions for Maximum Dose
3.4 Predictability of the Exceedance Function

4.0 ASSESSING RELIABILITY OF CURRENT KNOWLEDGE OF PARAMETERS
4.1 Measurement
4.2 Expert Opinion

5.0 APPLICATION OF A METHOD FOR PROBABILISTIC CALCULATIONS OF DOSE TO AN EXISTING COMPUTER MODEL

6.0 SOME CONCLUDING REMARKS

7.0 AREAS FOR FURTHER INVESTIGATION

REFERENCES

APPENDIX 1: APL Programs for Display of Empirical Functions and Sensitivity Analysis

APPENDIX 2: Notes Concerning the Beta Distribution

APPENDIX 3: Bounds on Probabilities via Chebychev's Inequality

EXECUTIVE SUMMARY

Mathematical modelling techniques are being used to predict the behaviour and consequences of various radioactive waste disposal strategies. Other mathematical techniques are being used to assess the quality of these predictions. It is essential that government regulators be equipped to judge the acceptability of these assessments. In this report we compare and evaluate some of the methods for assessing some aspects of the reliability of results obtained using mathematical modelling techniques.

The following notes outline the contents of the report by section.

1. Introduction

We indicate the necessity of using mathematical modelling methods for predicting the long-term behaviour and effects of radioactive waste disposal systems. A conceptual framework is provided for discussions of reliability of an individual mathematical model: the dependability of information concerning its parameters, the reliability of performance scenarios, and model quality. We discuss the role of expert opinion in assessing data reliability and the reliability of results obtained using modelling. A simple mathematical framework is defined for subsequent discussions of various methods for assessment of prediction reliability. We outline various ways of specifying parameter values and their reliabilities: point estimates and auxiliary statistics, confidence intervals, and probability distributions on the ranges of parameters. We introduce methods for converting information about the reliabilities of parameters into information about the reliabilities of predictions: interval arithmetic, Monte Carlo methods, "analytical methods".

2. Approaches to the Study of a Mathematical Model's Sensitivity

We introduce the notion that uncertainties about data propagate through a mathematical model to result in uncertainties about the model's output. The basic ideas behind the Monte Carlo method are set out. We record some fundamental definitions necessary for a discussion of statistical inference as it relates to applications of Monte Carlo method in the present context: probability density functions, cumulative distribution functions, exceedance functions, and the probability integral transformation. We indicate the

importance of exceedance functions in the study of dose limit exceedances. Some representative, alternative sampling plans are introduced for use in Monte Carlo sampling: completely random sampling, Latin hypercube sampling, and full factorial sampling. We describe their role in Monte Carlo experiments, and compare their properties. T-statistics, which are of special importance in Latin hypercube sampling, are introduced. We present some of the more important properties of T-statistics, and other statistics, under each of the three sampling schemes. Application of the so-called adjoint method to the kinds of problems that have, hitherto, been treated using Monte Carlo methods is outlined. We make a comparison of results obtainable using the adjoint and Monte Carlo methods, and we venture to caution regulators concerning one or two points.

3. Assessing the Exceedance Function for Dose

We suggest the exceedance functions's appropriateness in discussing the probability of exceeding a given dose level. Definitions of exceedance function estimators are presented for completely random and Latin hypercube sampling. We review the availability of ancillary statistics for use in conjunction with these estimators, and we outline the properties of these estimators. A method for assessing the sensitivity of an estimator of the exceedance function to perturbations of model parameter distributions is presented. We offer general comments concerning the application of reliability of assessment methods.

4. Quality of Current Knowledge Relating to Models of Ecosystems Subject to Discharges from Uranium Mill Tailings

We introduce the problem of assessing expert opinion in terms of probability distributions, and we describe the assessment method used to obtain the results presented later in the report.

5. Application of a Method for Probabilistic Calculation of Dose to an Existing Model

An existing mathematical model is described briefly, and we outline how the computer code which embodies the model was modified in order to make a Monte Carlo study of the model using Latin hypercube sampling. We also describe two small, auxiliary

programs for displaying exceedance and probability density functions.

The opinions of three experts concerning various aspects of the reliability of the existing model are presented, both in written form and in terms of probability density functions. We present an application of techniques based on Latin hypercube sampling to make a preliminary examination of the sensitivity of model outputs to perturbations of parameter probability distributions.

6. Some Concluding Remarks

We offer general comments concerning the usefulness and validity of results of methods for assessing modelling results reliability.

7. Areas for Further Investigation

We suggest other investigations which will yield useful information for regulators.

1.0 INTRODUCTION

1.1 Background

Above-ground tailings piles have been used for disposal of uranium mill wastes for many years. Uranium mine tailings potentially impact the environment in a variety of ways. The policies applied in governing their design and management have important consequences for all who share the environment.

The performance of any uranium mill tailings management area must be assessed over a very long period — of the order of 10,000 years. As a consequence, neither site monitoring nor experimentation can be soley relied upon for long term performance assessment. Recourse must be made to predictive modelling.

This study concerns itself with the behaviour of a particular mathematical model which simulates an aquatic ecosystem into which uranium mill tailings seepage flows. The model is that described in IEC BEAK (82).

The mathematical model takes the form of a computer program which can be used to calculate summary dose predictions from which tailings management and design inferences can be made. Since the potential consequences of decisions made on the basis of results from this computer model, or any other like it, are of such importance, considerable interest attaches to the reliability of model prediction. The present study concerns itself with an examination of methods for considering the reliability of predictions from this class of models.

1.2 General Prediction Reliability Considerations

We adopt the conceptual framework suggested by Cranwell and Helton (81) in treating the reliability of model predictions. There are three primary components to model prediction reliability:

- the fidelity with which the model emulates the system to which it corresponds;
- the dependability of data about model parameters;

- the dependability of knowledge of system performance scenarios that are embedded in the model.

The fidelity with which a model emulates a system depends upon factors which include:

- the successful selection and definition of the most salient or significant features and processes of the system for modelling;
- the accurate representation of system features and processes in terms of mathematical or computational analogues;
- appropriate parameterization of mathematical or computational models;
- accuracy and applicability of numerical techniques applied in conjunction with mathematical formulation of models.

Since this study takes the mathematical model for study as a given, the component of reliability relating to model quality is not further discussed in this report.

There are a number of components to system performance scenario dependability:

- scenario completeness;
- accurate identification and definition of parameters which describe scenarios;
- knowledge of frequency or probability of scenario event occurrences.

Over most of the time span covered by a complete scenario the processes in the system act in more or less continuous fashion. However, it must be assumed that both major and minor disruptions will occur and that they will alter conditions under which the normal processes operate, or even change the states of the processes. This means that a complete description of a system must take into account how discrete, perhaps disruptive, events will effectively re-initiate continuous processes, as well as what effect these events themselves will have.

The IEC BEAK model simulates a system of continuous processes, but does not provide for representation of the discontinuities which figure in any complete scenario, or collection of scenarios. However, we review some possible scenario features which

should receive further attention, in Section 4.1. We also offer comments on applicability of available methods to the study of prediction reliability with respect to scenarios, in Section 2.3.

The third component of model prediction reliability has to do with input data reliability:

- selection of appropriate values for model input;
- data interpretation and possible misuse of data;
- variation of data.

No objective characterization of all important sources and levels of data uncertainty exists for the aquatic ecosystem modelled by the IEC BEAK system, or for any other complicated systems. But, information about data uncertainty is available in the form of expert opinion. This report indicates means by which such expert opinion can be ellicited and summarized, and then used in studying the mathematical model of interest. We also provide a summary of existing knowledge relating to parameters in the model under study.

This report concerns itself with methods which may be used for characterizing the reliability of a model's predictions in numerical terms.

1.3 Characterizing Prediction Reliability in Numerical Terms

Prediction reliability analysis should be distinguished from certain forms of "sensitivity analysis". In these sensitivity analyses the model's responses to arbitrary perturbations of its parameters are studied. Since the sizes of these perturbations are not deliberately scaled to approximate variations ascribable to actual parameter uncertainty this sensitivity analysis tells more about the structure of the model than about prediction error. As Gardner et al (80) remark, the "situation is particularly dangerous for parameters .....which have small sensitivity coefficients but can dominate prediction error...as a result of measurement uncertainty" — a sensitivity coefficient for a model parameter may be small, but, against this, the parameter's value may be known only imprecisely, and data uncertainly will then overwhelm low parameter sensitivity. Clearly, it is important to determine how data and scenario reliabilities influence prediction reliabilities.

The mathematical model described in anonymous (82) gives an output vector of results as a function of time, mediated by a collection of model parameters that specify characteristics of the system that is being modelled. If we let H represent this mathematical model then we have

$$\underline{y} = H(t;\underline{x})$$

where $\underline{y}$ is a vector of output results,
t is time,
and $\underline{x}$ is a vector of model parameters.

The reliabilities of the predictions that the model generates depend in part on the reliability of the model parameters. In fact, numerous authors speak of parameter uncertainties that "propagate through" a model to manifest themselves as output uncertainties (Dormuth and Quick (80), Gardner et al (80), etc.).

We are therefore concerned in this report with methods for characterizing data dependability, and, having done this, with determining how output reliability can be assessed, particularly as output reliability arises as a reflection of data dependability. Such determinations are difficult, mainly because realistic models of ecosystems are so complicated; resulting in straightforward analytical methods being typically unavailable.

There are a number of ways to convey information about data dependability where values are given for data items. Data items can be given as point estimates with information about possible dispersion about the point estimate, they can be given as interval estimates, or they can be expressed as probability density functions.

When a parameter value is given as a point estimate there is the temptation to use a value which reflects the "worst case". However, as Dormuth and Quick (80) point out, this merely side-steps the problem of parameter variability: results produced using worst-case values tend to induce researchers to attend to situations which are of high consequence and to ignore other significant ones which may be much more likely. Point estimates which may be "most likely", or "most conservative", might be employed but they suffer from the same basic flaw.

A point estimate of a parameter should be accompanied by one or more values which indicate the dependability of the point estimate. The selection of appropriate values that indicate dependability will depend upon knowledge of, and/or assumptions about, the distribution on the point estimator involved. For example, it would be appropriate to select a standard error to express the dependability of a point estimate known to have come from a normal population. This is to say that correct specification of a point estimate, accompanied by a measure of its dependability, amounts to the specification of the distribution on the point estimator (more or less). Information about parameters is rarely available in this form.

Information about model parameter values, and the dependability of them, can also be conveyed using so-called confidence intervals[1]. The main drawback of this mode of expression is that a confidence interval does not, in itself, convey any information about the relative likelihoods that the value of the parameter represented will fall within various subsets of the confidence interval. Usually parameter values at the extremes of a confidence interval are less likely to occur than those toward the centre of its range, but this fact is not conveyed by the confidence interval.

The most straightforward way in which to express uncertainty about parameters is via probability distributions on them, which is the last of the three means mentioned by

[1] confidence interval: Let $\hat{r}$ be a statistical estimator of a model parameter r, based on random samples of r of the form $R_1, R_2, R_3, \ldots R_k$: that is, $\hat{r} = (R_1, R_2, R_3, \ldots R_k)$. Let us assume that one has grounds for asserting that

$$\Pr\{\hat{r} - c < r < \hat{r} + c\} = p,$$

where c is some constant. That is, the probability that the random interval $(\hat{r} - c, \hat{r} + c)$ includes the "true" value is p.

Now, let $r_1, r_2, r_3, \ldots r_k$ be an actual sample of values of r and let $r' = \hat{r}(r_1, r_2, \ldots, r_k)$. Then the interval $(r' - c, r' + c)$ is said to be a "100p per cent confidence interval for r." Beware of statements such as

$$\text{"}\Pr\{r' - c < r < r' + c\} = p\text{"}.$$

Since $r' - c$ and $r' + c$ are constants this statement is meaningless. In fact, some critics are more troubled by aspects of the concept of confidence intervals related to this type of statement than they will ever be about the "drawback" mentioned above.

Cranwell and Helton (81). Information about model parameters is seldom available in this form but fortunately methods exist for helping experts to express their knowledge of parameters in this form.

Once data and scenario reliabilities have been expressed in numerical terms it is necessary to apply some method for translating these reliabilities. There are a number of classes of these methods: there is interval arithmetic, there are Monte Carlo methods, and there are analytical methods.

Interval arithmetic: Suppose that a variable, z, is given as the sum of two other variables, x and y; that is, z=x+y. Suppose further that x is known to assume a value somewhere on the interval (x_{lower}, x_{upper}), and that y is known to have a value on the interval (y_{lower}, y_{upper}). Then z has a value on the interval $(x_{lower}+y_{lower}, x_{upper}+y_{upper})$. This implicitly defines an arithmetic operator which we can call interval addition, and denote by $\bullet$. $(a,b) \bullet (c,d) = (a+c,b+d)$. Similarly, one can define interval multiplication, interval exponentiation, etc., and the collection of rules for operating on interval quantities is called interval arithmetic.

Interval arithmetic is occasionally used in studying the performance of computer codes for numerical computations. An existing computer code that performs a numerical computation is processed using special software which translates ordinary arithmetic operations into interval arithmetic operations. Then, when the modified code is executed it performs all the calculations in the code using interval arithmetic so that the propagation of errors can be studied.

Clearly, interval arithmetic could be used in studying uncertainty propagation, where data reliability is given in terms of confidence intervals. However, as mentioned earlier, confidence intervals are not well suited to applications involving probabilistic statements about reliability, and, hence, interval arithmetic is not discussed further in this report. See Kulisch and Miranker (81) for an introduction to the subject of interval arithmetic.

Monte Carlo Methods: Monte Carlo methods are naturally suited, and directly applicable, to the study of reliability propagation in situations in which data reliability is expressed in terms of probability distributions. These methods may also be used in situations in which data reliability is given in terms of point estimators from known

distributions, where characteristics of the distributions other than their locations are given in terms of adjunct collections of numbers. Monte Carlo methods are treated more extensively below.

Analytical Methods: There are many methods for studying the sensitivity of a model in terms of its partial derivatives. Among these, the adjoint method distinguishes itself by its ability to provide probabilistic information about a model's response to data reliability. The adjoint method is applicable where data reliability is given in terms of variances, or, perhaps, certain other measures. The applicability of the adjoint method is considered more extensively below.

2.0 APPROACHES TO THE STUDY OF A MATHEMATICAL MODEL'S SENSITIVITY

2.1 Monte Carlo Methods

For a given time, t, a mathematical model H (.;.) transforms vectors of model input parameters into vectors of outputs. This implies that, if reliability information about the model parameters is given in terms of probability distributions over the parameters, then a model H(.;.) induces probability distributions over the outputs, by transformation. This is really just a probabilistic formulation of the idea that errors "propagate through" a model. For any model which is complicated enough to represent reality it will be impossible to calculate the mathematical forms of the distributions on the outputs, given the mathematical forms of the model and of the distributions on the inputs. Monte Carlo methods are some alternative approaches to obtaining information about distributions on outputs for various input distributions. These methods arise quite directly from the above probabilistic formulation of error propagation.

In essence,the Monte Carlo approach involves drawing some large number of samples from the input distribution, transforming each element of this sample using the mathematical model to form a sample from the output population, and then applying statistical method to this output sample to characterize the distribution on the output population. There are many alternative ways of organizing the input distribution sampling in a Monte Carlo experiment, and there are many alternative ways of making statistical inferences from the output sample. The statistical properties of these alternatives vary widely, and, as a consequence, so do their suitabilities for various applications.

2.1.1 Fundamental Preliminaries

This section serves only to define a few fundamental terms and ideas, including the probability integral transformation. (For an introduction, see Heathcote (71) for example).

Figure 2.1 displays a typical probability density function (pdf) defined over the range of a real-valued, scalar random variable. The area under the curve is 1, reflecting the fact that the probability of all possible single realizations of the random variable to which the

function corresponds sum to one. The area under the curve between any two distinct ordinates is the probability that the random variable will assume a value somewhere between these two ordinate values upon any given realization of it.

The function that gives the probability that a random variable, X, does not exceed a given value, x, is called the cumulative distribution function (cdf). (This function gives the left cumulative values of the area under the pdf). The cdf corresponding to the pdf in Figure 2.1 is depicted in Figure 2.2. Formally, if f(x), $-\infty < x < \infty$, is a pdf then the corresponding cdf is

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Since the area under the pdf is 1, $F(\infty) = 1$, or, in other words, the range of the function F(.) is (0,1).

Most of us seem to find it easier to think in terms of probability density functions — and often the cumulative distribution function is defined only for convenience of exposition, as is the case in this report.

The inverse of the cdf shown in Figure 2.2 appears in Figure 2.3. Formally, the inverse of a cdf F(.) is given by

$$F^{-1}(y) = x \text{ such that } y = F(x)^*$$

$F^{-1}(.)$, like the cdf, is defined only for convenience of exposition. (In "geometrical" terms $F^{-1}(.)$ is just the reflection of F(.) in the line y = x.)

A function which is of greater intrinsic significance, in the context of this report, is the so-called exceedance function. The exceedance function gives the probability that the value of a random variable, X, will be greater than a given x. It is given by

$$S(x) = 1 - F(x) = \int_{x}^{\infty} f(t)dt.$$

* We assume here for simplicity that x in this relation is uniquely defined.

Figure 2.1
Typical
Probability density function (pdf)

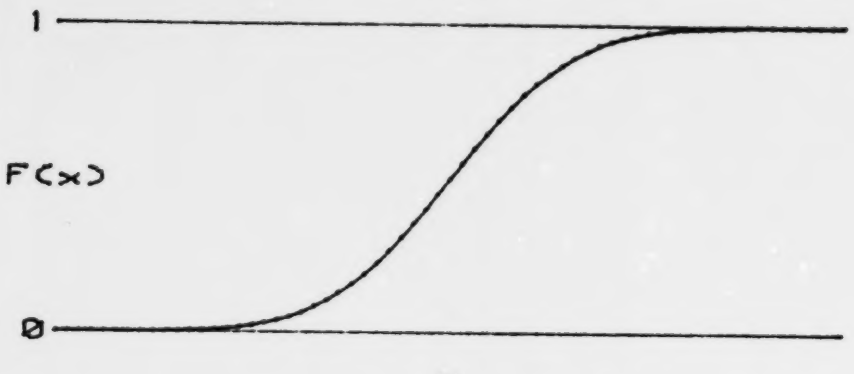


Figure 2.2
Cumulative distn function (cdf)
corresponding to pdf
shown in Figure 2.1

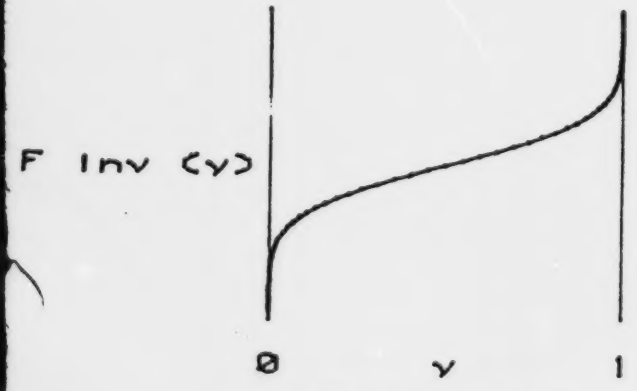


Figure 2.3
Inverse of
cumulative distribution function

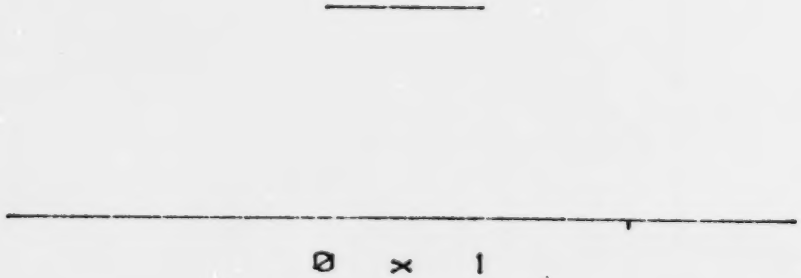


Figure 2.4
Uniform pdf on the unit interval

In this report we are concerned with the reliability of estimates of the probability that dose exceeds a given value. Thus we are studying the exceedance function for dose.

If a random variable U is uniformly distributed on the unit interval[1] as shown in Figure 2.4 then the random variable $X = F^{-1}(U)$ is distributed according to the pdf f(.) (or, alternatively, according to the cdf F(.), since the pdf and cdf are alternate representations of the "same thing"). In other words random deviates for an arbitrary population can be generated by applying the transformation $F^{-1}(.)$ to random deviates generated from a uniform distribution on the unit interval.

We apply this principle in describing how random deviates are sampled under the completely random and Latin hypercube sampling schemes in Section 2.3.

2.1.2 Some Sampling Plans

For discussion purposes, we describe three Monte Carlo sampling plans: completely random sampling (often called simply, random sampling), Latin hypercube sampling, and what we term complete factorial sampling.

In order to describe the three sampling procedures we adopt the following notation, in which it is assumed that a Monte Carlo study of a model with K parameters will involve sampling N times from the distribution(s) on the K parameters. Let $X_k : k=1(1)K$ be the set of model parameters. Let $F_K : k=1(1)K$ be the set of cumulative distribution functions corresponding to these parameters. That is, $F_k(.)$ describes how the probability of the occurrence of values of X_k varies over the range of X_k. $F_k^{-1}(.) : k=1(1)K$ is the set of inverse cumulative distribution functions corresponding to the X_k.

Completely Random Sampling: This is the simplest method to apply. It is of little practical use; this description is included only for comparison purposes.

[1] A random variable is "uniformly distributed on the unit interval" if at any realization of the random variable all values between 0 and 1 (that is, all values in the unit interval) are equally likely to occur.

The following algorithm indicates how to choose a completely random sample of N realizations of the vector of K input variables for a model:

Step 1: assign 1 to n
Step 2: assign 1 to k
Step 3: generate one uniform random number on the unit interval; call it U
Step 4: calculate $X_k=F_k^{-1}(U)$
Step 5: increment k by 1
Step 6: if k does not exceed K then go back to step 3

(Apply the mathematical model to the input value $X_1, X_2, X_3, \ldots, X_k$ at this point in the algorithm to obtain one model output y_i.)

Step 7: increment n by 1
Step 8: if n does not exceed N then go back to Step 2

Steps 2 through 6 result in the generation of one collection of model inputs, $X_1, X_2, X_3, \ldots, X_k$. Note that the determination of a new value for an X_k does not depend on what values have been chosen for any preceeding X_k's. This means that over the course of the Monte Carlo run, some regions of the range of values of each X_k will most likely be over-represented (and others under-represented).

Latin hypercube sampling: This is a relatively new technique which seems to have been described first by McKay et al (79).

For reasons of simplicity of exposition the following description of the method applies to the case where the X_k are (probabilistically speaking) independent, and the X_k are real-valued. Accommodations to situations where these conditions do not obtain are for the most part easily achieved.

Let R_k be the range of possibility values for the parameter X_k. Let $I_{k,1}, I_{k,2}, \ldots, I_{k,N}$ be a set of non-overlapping intervals which when taken together cover R_k. (For reasons to be outlined later the set $I_{k,1}, I_{k,2}, \ldots, I_{k,N}$ is usually chosen so that the probabilities $\Pr\{X_k \in I_{k,n}\}$, n=1(1)N, are within an order of magnitude of one another).

Under Latin hypercube sampling exactly one selection for variable X_k is made from each of the subsets of R_k given by $I_{k,1}$, $I_{k,2}$, ..., $I_{k,N}$, over the course of the N Monte Carlo trials. Thus, all parts of the range of each X_k are represented somewhere in the samples generated during a Monte Carlo run. This characteristic bears important consequences which are discussed later in this report.

In order to represent an algorithm for Latin hypercube sampling we use a vector of lists called unused to record which parts of the range of each input variable have yet to be sampled. Thus unused has the following initial value.

$$\underline{\text{unused}} = [(\text{unused})_k] = [(1,2,3,\ldots,N)_k],\ k=1(1)K.$$

We also use two matrices called lo and hi which are defined in the following way:

$$\underline{\text{lo}} = [(\text{lo})_{k,n}] = [F_k(\inf I_{k,n}))_{k,n}]$$
$$\underline{\text{hi}} = [(\text{hi})_{k,n}] = [F_k(\sup I_{k,n}))_{k,n}] :\ k=1(1)K,\ n=1(1)N.$$

inf (short for "infimum") $I_{k,n}$ is the left limit of the interval $I_{k,n}$. Thus, $F_k(\inf I_{k,n})$ is the left limit of the nth range of values for the kth parameter. This means that lo is a matrix of lower limits of intervals of values which are to be sampled from to provide values for parameters. hi is the corresponding matrix of the upper limits of these intervals (sup being sort for "supremum"). In other words lo and hi define collections of intervals which are the images of the intervals of the ranges of the model input parameters under the distribution functions $\{F_1(.), F_2(.), \ldots, F_k(.)\}$.

When a random deviate distributed according to distribution function $F_k(.)$ is to be selected from the interval $I_{k,n}$ the method outlined in Section 2.1.1 is used. That is:

- a deviate from the uniform distribution on the unit interval is selected
- a linear transformation is applied to this deviate so that it is as if the result had been drawn from the interval $(F_k(\inf I_{k,n}), F_k(\sup I_{k,n}))$ (which is where the matrices lo and hi are employed)
- this deviate, call it U, is subjected to the transformation $X=F_k^{-1}(U)$.

X is the desired random deviate.

The following steps constitute an algorithm for Latin hypercube sampling.

Step 1:	assign 1 to n
Step 2:	assign 1 to k
Step 3:	make a random selection from the list unused[k] and call it p
Step 4:	delete p from the list unused[k]
Step 5:	generate a uniform random deviate in the interval (lo[k,p], hi[k,p]), and call it U
Step 6:	calculate $X_k = F_k^{-1}(U)$
Step 7:	increment k by 1
Step 8:	if k does not exceed K then go back to Step 3

(Apply the mathematical model to the input value $X_1, X_2, X_3, \ldots, X_k$ at this point in the algorithm, to obtain one model output y_i.)

Step 9:	increment n by 1
Step 10:	if n does not exceed N then go back to Step 2

Steps 2 through 8 result in the generation of one collection of model inputs, $X_1, X_2, X_3, \ldots, X_k$. As mentioned before unused contains lists of numbers of intervals from which samples have yet to be taken. These numbers are deleted from the lists in Step 4 immediately after the corresponding intervals are sampled. Thus each of the N intervals for each of the K variables is sampled exactly once.

Full Factorial Sampling: If R_k is the range of possible values for the parameter X_k then let $\{I_{k,1}, I_{k,2}, \ldots, I_{k,M}\}$ be a set of non-overlapping intervals which cover R_k, taken together. (Note that the number of intervals in the above set, M, is not necessarily the same as the number of samples to be generated, N. Under Latin hypercube sampling the number of intervals in each range partition is the same as the number of samples to be drawn).

Under full factorial sampling, to produce one set of parameter values, $X_1, X_2, X_3, \ldots,$

X_K, one combination of intervals is generated by selecting one of the intervals for each variable, and, over the course of the Monte Carlo experiment the M intervals for the K parameters are combined in every possible way so that M^K combinations of intervals are selected. For example, if M=2 and K=3 then the following collections of intervals are generated over the course of a Monte Carlo experiment.

sample	intervals		
1	$I_{1,1}$	$I_{2,1}$	$I_{3,1}$
2	$I_{1,1}$	$I_{2,1}$	$I_{3,2}$
3	$I_{1,1}$	$I_{2,2}$	$I_{3,1}$
4	$I_{1,1}$	$I_{2,2}$	$I_{3,2}$
5	$I_{1,2}$	$I_{2,1}$	$I_{3,1}$
6	$I_{1,2}$	$I_{2,1}$	$I_{3,2}$
7	$I_{1,2}$	$I_{2,2}$	$I_{3,1}$
8	$I_{1,2}$	$I_{2,2}$	$I_{3,2}$

In the first sample a value for model variable 1 would be selected from interval $I_{1,1}$, a value for variable 2 from $I_{2,1}$, and for variable 3 from $I_{3,1}$.

In order to represent an algorithm for full factorial sampling we use two matrices called lo and hi, which are analogous to those defined for Latin hypercube sampling:

$$\underline{lo} = (lo)_{k,m} = F_k(\inf I_{k,m}))_{k,m}$$
$$\underline{hi} = (hi)_{k,m} = F_k(\sup I_{k,m}))_{k,m} \cdot \quad k=1(1)K,\ m=1(1)M.$$

As before, lo and hi define collections of intervals which are the images of the intervals $I_{k,1}$, $I_{k,2}$, $I_{k,3}$, ..., $I_{k,m}$, k=1(1)K, under the distribution functions $F_1(.)$, $F_2(.)$, ..., $F_K(.)$. When an element, U, is drawn from a uniform distribution on $\left[\underline{lo}_{k,m}, \underline{hi}_{k,m}\right]$, and the transformation $X=F_k^{-1}(U)$ is applied, the result is a random number, X, from the parent population with distribution function $F_k(.)$, restricted to the subrange of variable k given by $I_{k,m}$.

The following steps constitute an algorithm for full factorial sampling.

Step 1: generate the (first or) next combination of variable range interval numbers $(n_1, n_2, ..., n_K)$ (where each $n_i \in \{1, 2, 3, ..., M\}$)

Step 2: assign 1 to k

Step 3: generate a uniform random deviate in the interval $(\underline{lo}[k,n_k], \underline{hi}[k,n_k])$, and call it U

Step 4: calculate $X_k = F_k^{-1}(U)$

Step 5: increment k by 1

Step 6: if k does not exceed K then go back to Step 3

(Apply the mathematical model to the input value $X_1, X_2, ..., X_K$ at this point in the algorithm, to obtain one model output y_i.)

Step 7: if there are more interval combinations to generate then go back to Step 1

2.1.3 Notes Concerning Inference Procedures

A Monte Carlo experiment results in the generation of a sample of model outputs from the population of all possible model outputs. Statistical method is used to make inferences about the population from the sample.

There are two broad categories of statistical inference methods: there are those, called parametric techniques, which may be applied when the probabilistic model of the population under study, and of the sampling scheme, are known up to a few parameters; there are also methods, called the nonparametric techniques, which are applicable in situations in which less is known about the population being sampled. In general, parametric techniques provide more reliable statistical inferences than do the corresponding non-parametric techniques, provided that the underlying distribution is known (and assuming, of course, that parametric techniques exist for the given distribution and sampling scheme). Few parametric methods are applicable to Monte Carlo model output samples, however, because it is usually impossible to deduce the form

of the distribution on the model outputs given the distributions on the inputs -- and a complicated model.

Example. Suppose that an automobile designer is investigating the body dimensions of the human beings who reside in North America in order to design the passenger compartment of a new car model. He decides that he needs to know, among other things, the fraction of the population's knee-heel distances that do not exceed 70 cm, and he has taken a "random" sample of 10,000 measurements of these distances on which he will base his inference.

If we assume for the moment that these distances are normally distributed (that is, that they have the Gaussian distribution) then one estimator of this fraction is

$$\hat{P}_{70} = \Phi\left((70-\bar{x})/\hat{s}\right)$$

where $\bar{x}$ and $\hat{s}$ are the sample mean and sample standard deviation for the sample of 10,000 measurements, and $\Phi(.)$ is the normal cumulative distribution function.

If we cannot assume an underlying normal distribution then an alternate estimator is

$\dot{P}_{70}$ = (number of individuals in sample whose knee-heel distance is not greater than 70) /10,000

If the population knee-heel distances really are normally distributed then $\hat{P}_{70}$ can be expected to perform somewhat better than $\dot{P}_{70}$. That is, "on average", it will provide an estimate which is closer to the true, population value than the estimate which $\dot{P}_{70}$ will provide. However, if the knee-heel distance population is distributed according to one of a certain class of distributions which are not normal then $\hat{P}_{70}$ can produce rather poor results.

There are many measures of the individual or comparative quality of statistical estimators. We define two of the simpler ones here, and offer comments.

- If $\hat{p}$ is an estimator of the population parameter p then the bias of $\hat{p}$ is the difference between p and the expected value of $\hat{p}$. $\hat{p}$ is said to be unbiased for p if the expected value of $\hat{p}$ is equal to p.

  If the distribution of $\hat{p}$ is not symmetrical about $\hat{p}$ then p imparts less useful information about p. Furthermore, there are numerous examples in the statistical literature of estimators that are unbiased but quite useless. Clearly, the property of unbiasedness must be interpreted with some caution.

- If $\hat{p}$ is an estimator of the population parameter p then the mean squared error of $\hat{p}$ is the expected value of the square on the difference between $\hat{p}$ and p. If $\hat{p}$ is unbiased for p then the mean squared error of $\hat{p}$ is the same as the variance of $\hat{p}$.

It must be borne in mind that in general the choice of an appropriate statistical method depends not only on the underlying population distribution (if it is known) but also on how elements of the population are sampled. (More about this in Sections 3.2 and 3.4). Notwithstanding this general principle, it is possible to define a large class of statistical estimators which may be applied to results from a variety of sampling procedures.

Again, if we let H represent a mathematical model then we have

$$\underline{y} = H(t;\underline{x})$$

where $\underline{y}$ is a vector of output results
t is time,
and $\underline{x}$ is a vector of model parameters.

For fixed time, t, let $y_1, y_2, y_3, \ldots, y_N$ denote a sample of scalar model outputs corresponding to the sample parameters $\underline{x}_1, \underline{x}_2, \underline{x}_3, \ldots, \underline{x}_N$. That is,

$y_n = y_n(t) = H(t;\underline{x}_N)$, n=1(1)N. (That is, t is suppressed in the notation for the model outputs.) Then one class of estimators for characterizing an output population is given by

$$T(y_1, y_2, y_3, \ldots, y_N; g) = \sum_{1}^{N} g(y_i)/N,$$

where g(.) is an arbitrary function. Depending upon one's choice of g(.), T(.;.) can be made to estimate any of a variety of characterizations of the output populations for fixed t. For example (McKay et al (79)), for g(y)=y, T(.;.) is the mean of the sample output parameters.

T(.;.) has a number of useful properties, depending on the sampling procedure used to obtain $y_1, y_2, y_3, \ldots, y_N$, and we provide further discussion later in this report.

2.1.4 Comparison of Techniques Associated with Monte Carlo Method, and Assessment of Monte Carlo Results

Consider studying a model involving K=2 input parameters using simple random sampling and Latin hypercube sampling, and assume for the moment that N=10 pairs of input parameters are deemed to be required to meet the level-of-precision requirements of the Monte Carlo study.

We will assume the reliabilities of each of the two parameters can be modelled by saying that the parameter is uniformly distributed on the unit interval.

To say that each parameter is uniformly distrubuted on the unit interval is to say that their density functions are

$$f_k(u) = \begin{cases} 1 & 0 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}, \quad k=1(1)2.$$

This implies that their distribution functions are

$$F_k(u) = \begin{cases} 0 & u < 0 \\ u & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}, \quad k=1(1)2.$$

We choose the same sets of intervals

$$
\begin{aligned}
I_{k,1} &= [0, .1) \\
I_{k,2} &= [.1, .2) \\
I_{k,3} &= [.2, .3) \\
&\ \ \vdots \\
I_{k,10} &= [.9, 1], \quad k=1(1)2,
\end{aligned}
$$

to cover the ranges of both of the input parameters for Latin hypercube sampling.

Then the following results seem typical of all those that might be obtained using completely random sampling and Latin hypercube sampling.

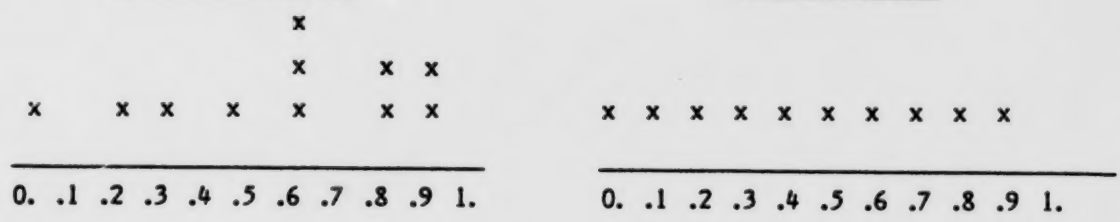
Samples

Completely Random		Latin hypercube	
.829	.801	.880	.638
.650	.379	.787	.962
.872	.868	.0578	.219
.956	.618	.510	.536
.0495	.578	.924	.146
.336	.198	.272	.841
.634	.0966	.302	.490
.680	.865	.608	.0895
.925	.236	.159	.393
.243	.457	.494	.719

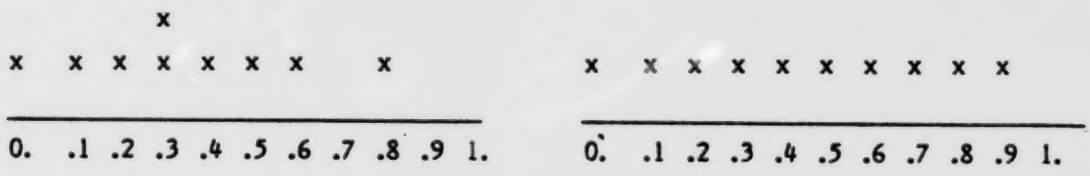
The following histograms point up an essential difference between the two results.

Histograms

First Variable



Second Variable



Even though each variable is uniformly distributed on the unit interval, the sample elements drawn by completely random sampling are "unevenly" spread over the range of possible values: two or three deviates fall in some categories and consequently other categories are unrepresented in the sample. In contrast, Latin hypercube sampling has generated a sample in which all of the intervals that the sampler selected in each variable's range are represented. It is because of this "stratified" sampling of each of the variables under study that Latin hypercube sampling can often achieve much higher efficiencies that can completely random sampling. This means that smaller samples of parameters need be drawn for the same model output estimate precision.

Consider now studying the same two-parameter models, but this time using full factorial sampling. For the sake of later comparison, let us assume that it is considered necessary

to ensure that representative values of each parameter must be selected from each of M=10 subranges of the parameters, where these subranges are defined as they were for Latin hypercube sampling. To simplify representation let us adopt labels for the intervals according to the following table:

interval	label
$I_{1,1} = I_{2,1} = [0, .1)$	"1"
$I_{1,2} = I_{2,2} = [.1, .2)$	"2"
$I_{1,3} = I_{2,3} = [.2, .3)$	"3"
.	
.	
.	
$I_{1,10} = I_{2,10} = [.9,1$	"10"

Then, under full factorial sampling the following collection of intervals would be generated:

Sample	intervals
1	1, 1
2	1, 2
3	1, 3
4	1, 4
5	1, 5
6	1, 6
7	1, 7
8	1, 8
9	1, 9
10	1, 10
11	2, 1
12	2, 2
13	2, 3
14	2, 4
15	2, 5
16	2, 6

17	2, 7
18	2, 8
19	2, 9
20	2, 10
21	3, 1
22	3, 2
23	3, 3
.	
.	
.	
89	9, 9
90	9, 10
91	10, 1
92	10, 2
93	10, 3
94	10, 4
95	10, 5
96	10, 6
97	10, 7
98	10, 8
99	10, 9
100	10, 10

Full factorial sampling results in $100=10^2$ ($=M^K$) collections of intervals.

In comparison the collection of intervals which was generated using Latin Hypercube sampling is as given in the following table:

sample	intervals
1	9, 7
2	8, 10
3	1, 3
4	6, 6
5	10, 2

6	3, 9
7	4, 5
8	7, 1
9	2, 4
10	5, 8

- Quoting from McKay et al (79):

  "One advantage of the Latin hypercube appears when the output y(t) is dominated by only a few of the components of $\left[\underline{x}\right]$.

  This method ensures that each of these components is represented in a fully stratified manner, no matter which components might turn out to be important."

  Note, however, that although one can arrange to ensure that all parts of the range of each model variable are represented in the Latin hypercube sample, Latin hypercube sampling results provide no information about mathematical interactions among model parameters. The model output for sample number 7 indicates results when parameter 1 is in interval number 4 and parameter 2 is in interval 5, but what would the results be if parameter 2 were in interval 6, with parameter 1 in interval 4? If we knew the model results for the latter possibility then we could determine the effect of changing parameter 2 from a value in interval 5 to a value in interval 6 while parameter 1 is in interval 4. However, Latin hypercube sampling provides no information about such interactions at all. Furthermore, it provides information only for randomly chosen combinations of model parameter values. It is unusual to know which combinations are most "interesting", but even if the interesting ones were known we could not make use of this information using Latin hypercube sampling.

- The primary purpose of this study is to demonstrate methods for measuring model output reliability as a function of model parameter reliability. We use probability distributions to describe reliability in the model parameters. Since these distributions cannot themselves be accepted unquestioningly it is important to entertain results which indicate the contribution of this latter uncertainty to uncertainty in the distributions on the model outputs.

Since Monte Carlo runs are so expensive to make it is usually not feasible to make many of them — with various distributions on the model parameters — in order to assess the effects of perturbing the parameter distributions.

This brings us to a significant virtue of Latin hypercube sampling: Although there are importants flaws in the technique Latin hypercube sampling does make it possible to assess the sensitivity of a mathematical model to perturbations of the probability distributions on model parameters — without making repeated Monte Carlo runs (Iman et al (80)). This is also possible with full factorial sampling but, because full factorial sampling may involve such exhaustive sampling of the model parameter space, it is likely to be prohibitively expensive. (It is not possible to obtain sensitivity analyses in this way using results from completely random sampling).

This sensitivity analysis approach is applied to estimation of a exceedance function from a Latin hypercube sample in Section 3.4.

- For Latin hypercube sampling the number of intervals into which the range of each model parameter is partitioned must be equal to the number of samples considered necessary to satisfy precision requirements in results.

This is in spite of the fact that some parameters might have little influence on model outputs and might therefore require less refined partitions. Although we have defined full factorial sampling to provide for the same number of subranges for each model parameter in this report, for the sake of simplicity of exposition, in fact the numbers of intervals in parameter partitions could all be distinct. That is, under full factorial sampling, each parameter's range needs only to be partitioned as finely as necessary; there are no artificial constraints as with Latin hypercube sampling.

Any Monte Carlo approach to model investigation has an essential weakness: model outputs are known only for a sample of parameter values, and, from such outputs, it is impossible to determine whether any or all regions of the mathematical function expressing the model have been adequately explored.

- In certain circumstances it may be advantageous to substitute a full factorial sampling scheme for a Latin hypercube sampling scheme. For example, suppose that 16 parameter set samples are required for a 2-parameter model. Under Latin hypercube sampling this implies that each parameter's range must be partitioned into 16 intervals. If, however, it is known that it is sufficient to partition each range into two intervals then a full factorial sampling plan could be used since $M^K=4^2=16$, which is the required number of parameter set samples.

- If T(.;.) is a (mean-) unbiased estimator under completely random sampling then T(.;.) is also a (mean-) unbiased under Latin hypercube sampling (McKay et al (79)). (That is, if the expected value, or mean, of T(.;.) under completely random sampling is a certain value, say t, then the expected value of T(.;.) under Latin hypercube sampling is t also).

- This implies that a large class of unbiased estimators exists for Latin hypercube sampling. A similar class can be defined for full factorial sampling.

- The variance of T(.;.) based on a Latin hypercube sample is less than the variance for a completely random sample, provided that H(.;.) (the mathematical model) is monotonic in each of the elements of the vector of model parameters x, and g(.) (in the definition of T(.;.)) is monotonic in its argument (McKay et al (79)). In fact, McKay et al demonstrate that estimates based on Latin hypercube sampling exhibit far less variance over a range of situations.

  In practical applications one of the difficult problems in applying Latin hypercube sampling may lie in determining whether the former monotonicity condition mentioned above is satisfied, and for what ranges of the model parameters. Even if this requirement is not satisfied, however, it may be possible to handle a problem in a piecewise fashion. The model parameter space is partitioned in such a way that the monotonicity condition holds over each subspace. Then a Monte Carlo simulation employing Latin hypercube sampling is conducted for each subspace.

- There are results in the literature that compare the precisions obtainable using completely random and Latin hypercube sampling, but there appear to be no general formulae which give the precisions of model outputs in numerical terms for

any of the three sampling schemes outlined above. This is a serious limitation in the use of Monte Carlo methods. It makes it more difficult to choose appropriate sample sizes, and more work must be expended in arranging to obtain the information needed to assess results. Owing to the complexities of mathematical models and sampling schemes there seems to be little likelihood of the appearance of such formulae. (It should go without saying that the report of any Monte Carlo experiment should include numerical information about the precisions of results obtained).

- The class of estimators defined in Section 2.1.3 includes many estimators whose statistical properties are well known for completely random sampling. Their properties are, for the most part, unknown for Latin hypercube and full factorial sampling. The decision whether to use completely random sampling, rather than one of the other types, would involve consideration of whether model parameter distribution sensitivity is the main goal, or improved knowledge of the statistical properties of estimates.

  We consider these questions further in Section 3.4.

- The Latin hypercube sampling algorithm, as described above, makes use of considerable computer storage. The arrays lo and hi each use 2N storage locations, and the array unused occupies KN storage locations. It should be mentioned, though, that these requirements can be considerably reduced, by keeping track of the intervals sampled for each parameter in an array of bits, and by generating the Latin hypercube parameter sub-range lists one at a time, by parameter, rather than all at once.

2.2 Analytical Methods

What we term "analytical methods" are those which relate model output variability to parameter and scenario reliability through partial derivatives of the model's response function with model parameters. A pre-eminent member of this class of techniques is that described by Thomas (82)[1] under the name adjoint method.

1 We are grateful to Dr. D.J. Martin of the AECB for dispatching a copy of this paper to us.

The adjoint method requires the somewhat tedious calculation of algebraic expressions for partial derivatives, but it seems a very efficient method for relating prediction reliability to data and scenario reliabilities, when data and scenario reliabilities are expressed in certain ways, and when prediction reliability is to be expressed in one of a certain collection of ways. In particular, if the reliabilities of the model parameters are given as variances about mean values, then the adjoint method can be used to provide variances on model predictions. Model prediction variances and means do not translate directly into probabilities of exceeding arbitrary limits but bounds on these probabilities can be obtained using Chebychev's Inequality,[1] for example. Furthermore, the sensitivity of these bounds to perturbations of model parameter reliabilities can also be assessed.

What the adjoint method does not appear able to provide, in its present state of development, are estimates of entire probability density functions, cumulative distribution functions, exceedance functions, and the like.

2.3 Summary Comment

Mathematical interactions among a model's parameters are undiscernible using Latin hypercube sampling; Thomas (82) argues this point forcefully and convincingly. We disagree, however, with his repeated statement that "the primary virtue of Latin hypercube sampling is the fact that it yields unbiased estimates of the probability density functions for computer outputs."[2] Although (mean-) unbiasedness is not a particularly compelling criterion, it is important to say that a "virtue" of Latin hypercube sampling is that it provides for unbiased estimation of an entire class of output distribution characterizations (not just of density functions), and that, furthermore, it makes it possible to treat the sensitivity of these estimates to input distribution perturbations.

1 See Appendix 3

2 Even this statement is not strictly true: the literature concerning Latin hypercube sampling proposes estimators for unbiased estimation of cumulative distribution functions rather than density functions. However, it does seem that some of the methods for density estimation from completely random samples can be adapted for use with Latin hypercube sampling.

Although Thomas sketches a mathematical treatment of output distribution moments as functions of parameter distribution moments, he does not indicate how one might obtain estimates of the sensitivities of various probabilities defined on outputs. We suspect that such mathematical treatments can be invented (we suggested one approach in Section 2.2). However, for the present, it seems that one is obliged to employ Monte Carlo methods to obtain estimates for objects such as probability density functions.

O'Neill et al (82) say:

> "Monte Carlo methods are ideally suited to the investigation of parameter uncertainty. They can account for all of the information available to the modeller and provide a straightforward estimation of resulting errors. Because it is impractical to inspect each iteration of the model, absurd as well as reasonable results are used to estimate means and variances of predicted values.....The purpose of this study is to explore possibile ways in which a priori information can be used to avoid physically impossible or unrealistic combinations of parameter values."

There are really two sides to this issue. A priori information should certainly be used to improve the reliability of the predictions that a model makes. On the other hand, though, there is a strong temptation to force decreased variation in a model's outputs by imposing artificial, more or less arbitrary, mathematical constraints and probabilistic covariance structure on the parameters of the mathematical model. The only way to detect such abuses would appear to be to appeal to disinterested parties to examine each mathematical model for the intrinsic validity of all of its components.

Although O'Neill et al speak only of Monte Carlo methods, these considerations should apply whether Monte Carlo or analytical methods are used for model study.

For the purposes of much of the remainder of this report, we have selected Latin hypercube sampling to represent the capabilities and limitations of Monte Carlo method in the context of the study of mathematical models.

3.0 ASSESSING THE EXCEEDANCE FUNCTION FOR DOSE

When the reliabilities of a mathematical model's inputs, and other factors, are themselves modelled using probability distributions the model can be thought of as transforming these distributions to yield outputs which have probability distributions defined over them. An expert's expression of his belief as a density function will usually be met with at least some incredulity, and, therefore, one must say that we are uncertain about these distributions. Since we are uncertain about the distributions on the inputs we must be uncertain about the distributions on the outputs. (More will be said about this issue in section 6.0).

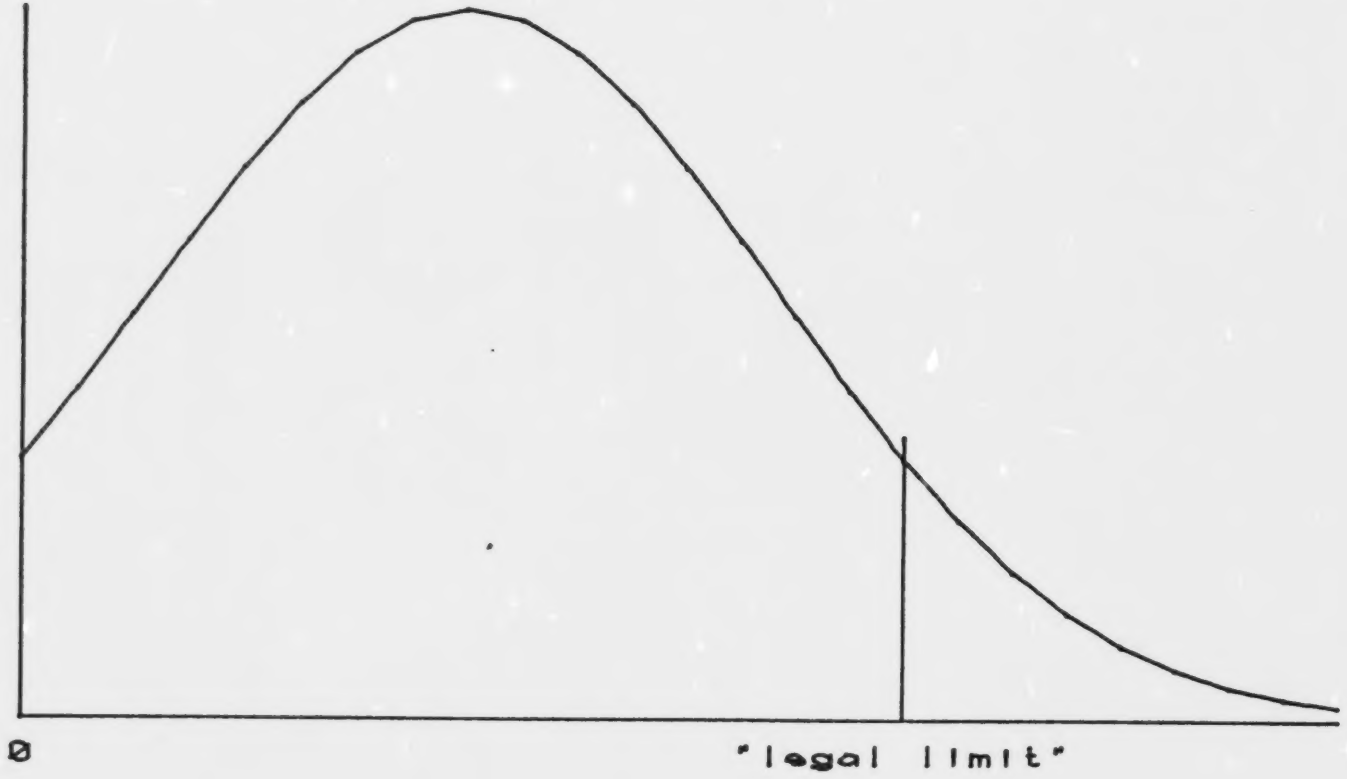
The central objective of this study is to review methods for measuring the reliability "associated with the calculated probability distribution of dose" (Coady (82)).

Figure 3.1 depicts a typical probability density function on "dose". Note that arbitrarily large doses, beyond any simply stated "legal limit" (say), can be expected to occur with non-zero probability. Therefore, attention naturally focuses on the area under the density function to the right of the legal limit. The name of the relation which gives the probability of occurrence of a random deviate greater than or equal to a given value is the exceedance function. The central objective of this study can be restated in terms of the exceedance function. It is to review methods for measuring the reliability associated with the exceedance function on dose.

3.1 Estimation of the Exceedance Function

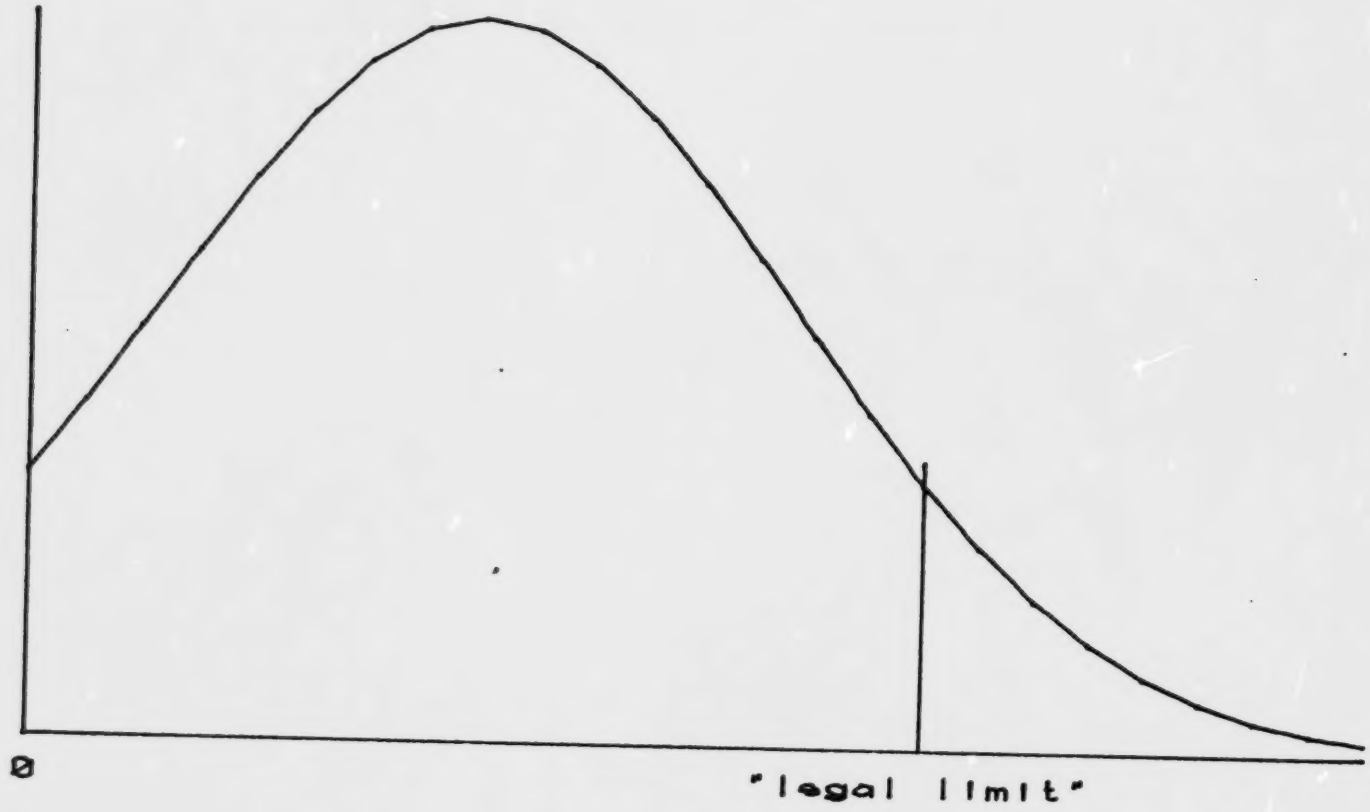
The exceedance function bears a close relationship to the cumulative distribution function. If the distribution function for a random variable X is F(.) then X's exceedance function is $S(x) = 1 - F(x)$. It is not surprising that most of the statistical theory concerned with distribution functions can be applied to exceedance functions.[1] We proceed now to define an estimator of the exceedance function which is analogous to the

1 For an introduction to this theory see Kendall and Stuart (73).



Typical density function on dose

Figure 3.1



Typical density function on dose

Figure 3.1

usual estimator of the distribution function. Let $X_1, X_2, \ldots, X_N$ be a simple, random sample[1] from a (continuous) population with exceedance function S(.). Let $(X_{(1)}, X_{(2)}, \ldots, X_{(N)})$ be the same set of values ordered by size so that $X_{(1)}$ is the smallest element of the original set, $X_{(2)}$ is the next smallest element after $X_{(1)}$, ..., and $X_{(N)}$ is the largest element of $X_1, X_2, \ldots, X_N$. Then an estimator of the exceedance function is the empirical exceedance function for a simple random sample.

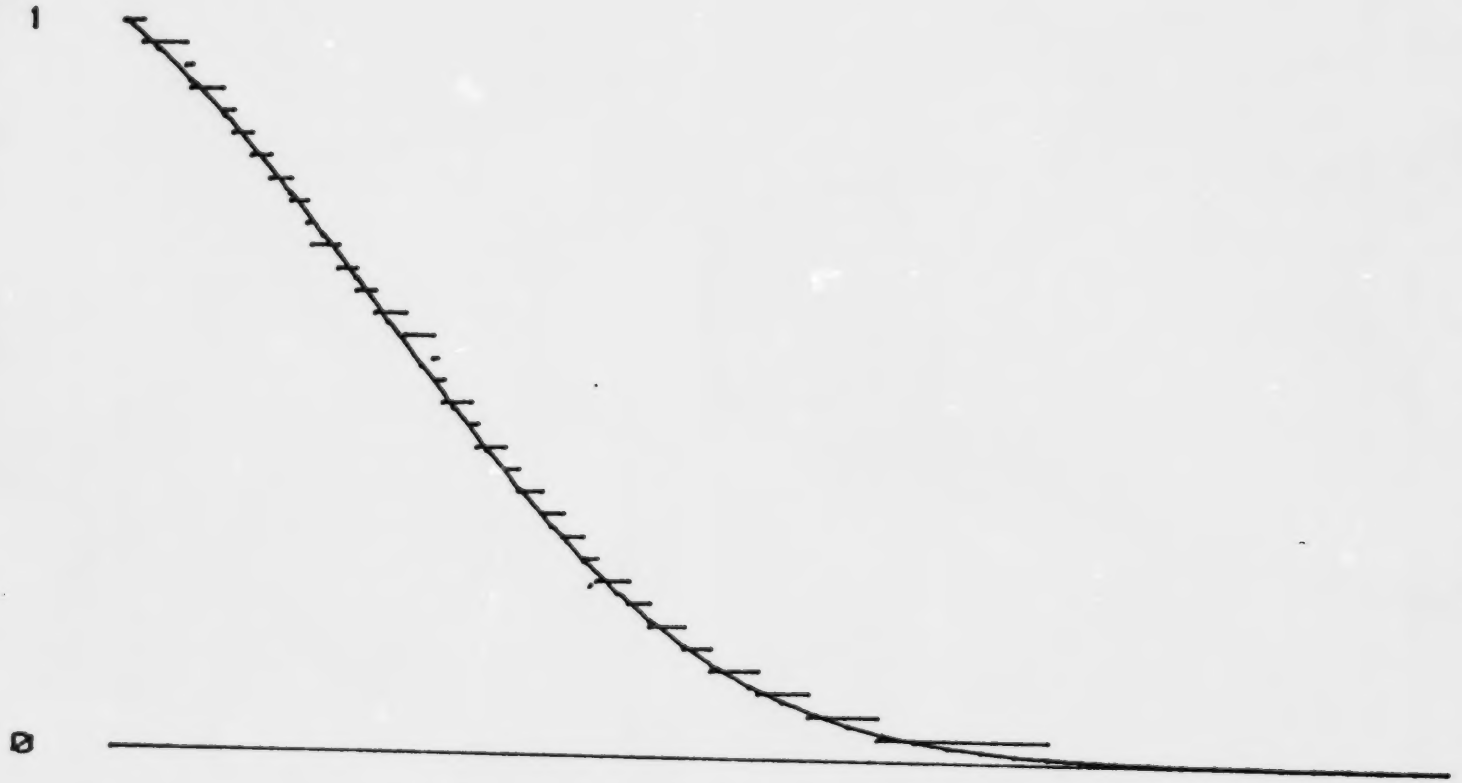
$$\hat{S}_{srs}(x) = \begin{cases} 1 & x < X_{(1)} \\ 1 - 1/N & X_{(1)} \leq x < X_{(2)} \\ 1 - 2/N & X_{(2)} \leq x < X_{(3)} \\ 1 - 3/N & X_{(3)} \leq x < X_{(4)} \\ & \cdot \\ 1 - \frac{n}{N} & \cdot \; X_{(n)} \leq x < X_{(n+1)} \\ & \cdot \\ & X_{(N-1)} \leq x < X_{(N)} \\ 0 & X_{(N)} \leq x \end{cases}$$

Figure 3.2 depicts an exceedance function, and an empirical exceedance function, constructed from a sample[2] from the population with that exceedance function. This exceedance function corresponds to the density function shown in Figure 3.1.

Note particularly how the empirical exceedance function approximates to the exceedance function.

1 $X_1, X_2, \ldots, X_N$ are said to constitute a simple, random sample if they are statistically independent, and are all distributed according to the survival function S(.).

2 The sample was -.93, -.77, -.75, -.63, -.59, -.52, -.45, -.37, -.31, -.29, -.19, -.12, -.05, .063, .17, .187, .21, .315, .34, .448, .5, .59, .67, .745, .8, .92, 1.03, 1.12, 1.23, 1.41, 1.6, 1.86, 2.5



Typical survival function (continuous)

Sample empirical survival function (steps)

Figure 3.2

Now let $X_1, X_2, \ldots, X_N$ be a sample generated by Latin hypercube sampling, and define $(X_{(1)}, X_{(2)}, \ldots, X_{(N)})$ as for simple, random sampling. An estimator for the exceedance function can be defined along the lines suggested by Iman and Conover (80).

$$\hat{S}_{lhs}(x) = \begin{cases} 1 & x < X_{(1)} \\ 1 - N^{K-1} p_{n_1} & X_{(1)} \leq x \leq X_{(2)} \\ 1 - N^{K-1}(p_{n_1} + p_{n_2}) & X_{(2)} \leq x < X_{(3)} \\ 1 - N^{K-1}(p_{n_1} + p_{n_2} + p_{n_3}) & X_{(3)} \leq x < X_{(4)} \\ 1 - N^{K-1}(p_{n_1} + p_{n_2} + p_{n_n}) & X(n) \leq x \leq X(n+1) \\ 1 - N^{K-1} \sum_{i=2}^{n} p_{n_i} & X_{(N)} > x \end{cases}$$

p_{n_i}, i=1(1)N, is 1/N times the probability in the interval from which $X_{(i)}$ was selected during Latin hypercube sampling.

There seems to be no explicit definition of an empirical distribution function for full factorial sampling in the literature.

3.2 Properties of the Empirical Exceedance function

For increasing sample size, the empirical exceedance function becomes an increasingly good approximation to the true exceedance function. The empirical exceedance function for a simple random sample is said to exhibit strong convergence to the exceedance function as sample size N increases.[1]

[1] That is, the probability that the maximum absolute difference between the value of the empirical survival function and the survival function over all x will exceed a given value can be made arbitrarily small by increasing the sample size.

Kolmogorov derived the distribution of the following statistic (see Kendall and Stuart (73))[1]:

$$K_N = \sup_x \left| \hat{S}_{srs}(x) - S(x) \right| .$$

The availability of this distribution means that confidence bounds can be easily constructed around a empirical exceedance function using published tables of the distribution.

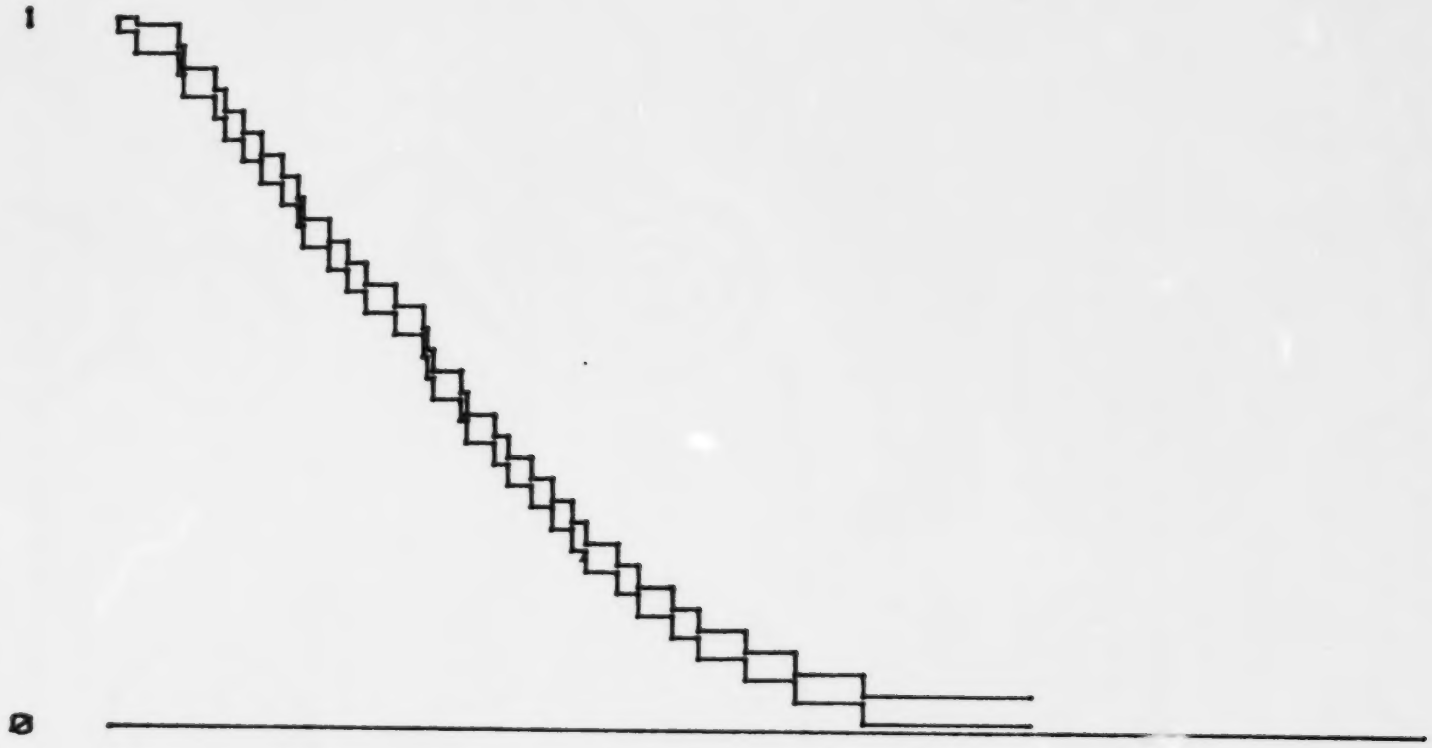
In rough terms, K_N is the maximum vertical distance between the true exceedance function S(.) and the estimated exceedance function $\hat{S}_{srs}(.)$. K_N is a measure of the closeness of $\hat{S}_{srs}(.)$ to S(.).

If k is a value such that $\Pr \{ K_N \leq k \} = p$, for some probability p, then a 100 p per cent confidence band can be constructed around $\hat{S}_{srs}(.)$ as the pair of step functions which are at most k units above and below $\hat{S}_{srs}(.)$ respectively. Figure 3.2b resembles 3.2a, with $\hat{S}_{srs}(.)$ replaced by confidence band for S(.) constructed using the Kolmogorov statistic K_N in this way. This confidence band covers S(.) with some arbitrary, high probability.

The width of the confidence band constructed using the Kolmogorov statistic is constant over the domain of the function. This is not a reflection of the true pattern of variability in the empirical survival function. In fact, in most cases, the variability is likely to be less at the extremes of the empirical exceedance function's domain. Confidence bands can be constructed using more sophisticated statistical theory, in such a way that the levels of variability over the empirical exceedance function's domain are more faithfully rendered. However, this example of confidence interval estimation should serve to illustrate the principle.

$S_{lhs}(.)$, on the other hand, does <u>not</u> exhibit strong convergence to S(.), and the distributions on any statistics such as K_N above seem to be wholly (and not surprisingly) unknown.

1 In actual fact, Kolmogorov derived the result for distribution functions, but it holds for survival functions.



Confidence band on survival function

Figure 3.2b

It can be shown that $\hat{S}_{lhs}(.)$ is unbiased[1], that is, that $E\left\{\hat{S}_{lhs}(x)\right\} = S(x)$ for each x. However, as can be seen from the definition of $S_{lhs}(.)$, if $\sum_{1}^{N} p_{n_i} < 1/N^{K-1}$ then $S_{lhs}(x) < 0$ for $x \geq X_{(N)}$. (This, in itself, means that $\hat{S}_{lhs}(.)$ cannot strongly converge.) This is one rationale for choosing parameter sub-ranges with equal probability. Another reason is presented in Section 3.4. The definition of $\hat{S}_{lhs}(.)$ can be adjusted to eliminate this characteristic, but the statistical grounds for doing so are almost lacking.

For any given x, S(x) gives the probability that the random variable characterized by S(.), X, will exceed x. Under simple random sampling, an estimate of this probability is given by p = n/N where n is the number of model outputs which exceed x, and N is the overall Monte Carlo sample size. An estimate of the standard error of this estimate is given by $\sqrt{p(1-p)/N}$. That is, the sampling error of this estimator is inversely proportional to the square root of the Monte Carlo sample size. An arbitrarily small standard error can be achieved by choosing N sufficiently large. (Or, for fixed N, the standard error measures the level of statistical sampling variation in the estimate of a proportion).

The estimator p can be applied to the results of Latin hypercube sampling. However, no estimate of the variance is available.

The size of the Monte Carlo sample of outputs required to produce satisfactory statistical inferences about the output populations depends on a variety of factors. The levels of variation in the model parameters, the exact nature of the mathematical model, and the levels of precision required in the model output characterizations are among these factors. We do not address the problem of sample size determination in this study. Throughout the rest of this report we assume that the sample size, always denoted by N, is a given.

3.3 The Exceedance Functions for Maximum Dose

The radiation dose rate experienced by a typical human individual is a non-constant

1 $\hat{S}_{lhs}(.)$ is a "T-statistic" (see Section 2) which is unbiased for completely random sampling. Hence it is unbiased for Latin hypercube sampling, as mentioned in Section 2.

function of time. One way to characterize the dose rate to a typical person is to use the maximum dose rate that such an individual experiences. The reliabilities of maximum dose rate predictions made by a mathematical model reflect the reliabilities of various aspects of the model. Thus, it is reasonable to speak of there being a probability distribution on maximum dose rate. Attention in this section now turns to estimators of the exceedance function, the estimator for simple random sampling, and that for Latin hypercube sampling.

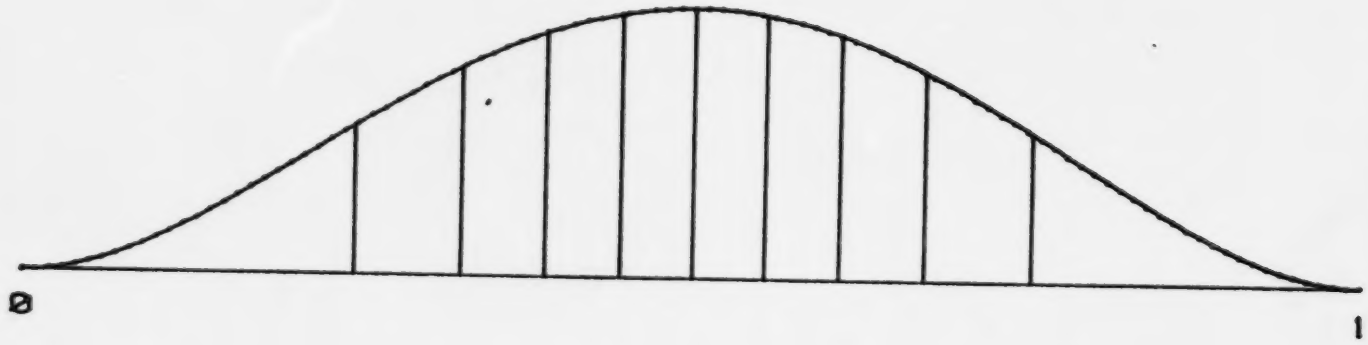
3.4 Dependability of the Exceedance Function

Although the statistical theory concerning empirical exceedance functions for simple random sampling is highly developed, it is not possible to investigate the sensitivity of such functions to parameter distribution variation without running and re-running the associated Monte Carlo experiment. On the other hand, there is little statistical theory applying to empirical exceedance functions for Latin hypercube sampling, but it is possible to investigate output distribution sensitivity using this method.

The argument supporting the idea that Latin hypercube sampling can be used in investigating the sensitivity of model output distributions to perturbations of the input distributions is a heuristic one - there appear to be no rigorous, mathematical proofs of this property of the method in the literature. We present the heuristic argument now.

Consider studying a five-parameter model, and consider the sampling of just one of the model's parameters under Latin hypercube sampling. For concreteness, let us suppose that ten samples of input parameters are to be selected, and that the distribution on the parameter in question is to be beta(3,3), as shown in Figure 3.3. Since ten samples are to be generated, the range of this distribution is partitioned into ten parts. We choose the partition such that the probability in each interval is 1/10, for reasons to be outlined presently. The vertical lines in Figure 3.3 partition the area under the density into ten equal parts.

Over the course of the Monte Carlo experiment, exactly one element will be selected from each of these subintervals, the model will be applied to all parameters and a sample of ten outputs will thus be generated. The estimator $\hat{S}_{lhs}(.)$, presented in Section 3.1, will be used to form an estimate of the exceedance function S(.).



Beta(3,3) density function

Figure 3.3

Since N=10 samples were generated and we have assumed that K=5 parameters are involved in the model, the probabilities p_{n_i}: i=1(1)n in $\hat{S}_{lhs}(.)$ are all equal to .01 = .1 x .1.

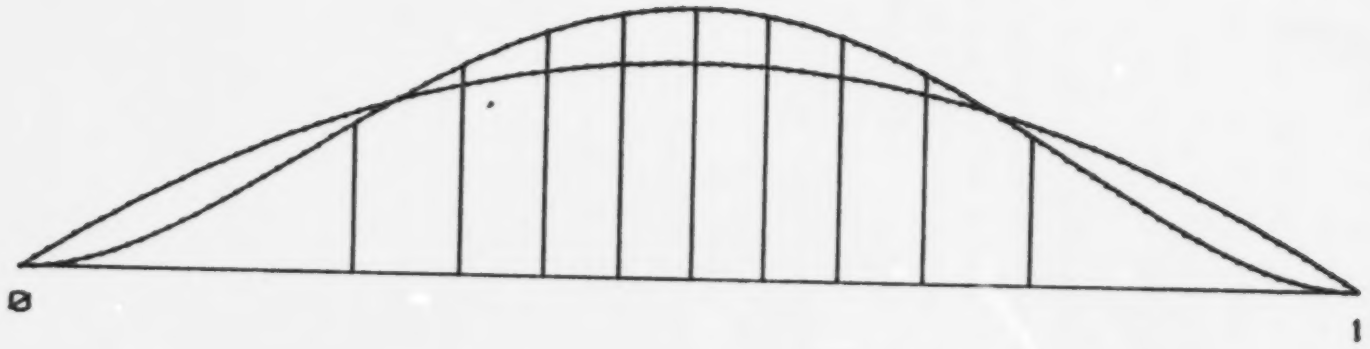
(Each parameter's range is to be partitioned into 10 equally probable intervals. That is, the probability of each of these subintervals is to be .1. The probability that a given subinterval will be selected in a particular sample is .1. Thus, the probability associated with a given subinterval in a given sample is .01 = .1 x .1.).

Suppose now that we require some indication of the sensitivity of $\hat{S}_{lhs}(.)$ to changes in the distribution on the parameter whose reliability has been summarized using the beta(3,3) distribution. Let us suppose, in fact, that this parameter's reliability might be better described using the beta(2,2) distribution, which is compared with the beta(3,3) in Figure 3.4. (We note that more probability mass is distributed away from the mean for the beta(2,2) distribution than for the beta(3,3).) This implies that we require knowledge of the effect on the model output distribution of an assumption of lower reliability concerning knowledge of the parameter.

Under the beta(2,2) distribution, the probabilities of the ten sub-ranges are not .01, as for the beta (3,3). They are .153, .098, .087, .081, .081, .081, .081, .087, .098 and .153 respectively. If, however, we had sampled from these same sub-ranges but from the beta(2,2) distribution, we would expect the elements chosen under Latin hypercube sampling to be "much the same". Thus, it is claimed that the results of having sampled from the beta(2,2) can be approximated by substituting the values .1 x (.153 .098 .087 .081 .081 .081 .081 .087 .098 .153) for the ten .01 values for the p_{n_i} in $\hat{S}_{lhs}(.)$.

Obviously this claim holds true, in the limit, as the chosen alternative, perturbed input distribution approaches the distribution actually used in Monte Carlo sampling. However, there appear to be no results which indicate how small the input distribution perturbations must be in order to obtain valid sensitivity results.

On intuitive grounds, the statement is also made that this approximate behaviour of Latin hypercube sampling is most likely to hold true when the sub-ranges are chosen to be of approximately equal probability. Under this constraint, perturbation of parameter distributions is likely to have the least effect, "on average", on the parameter values actually chosen. There are, apparently, no results which a designer can use in deciding



Comparison of sampling from

Beta(3,3) and Beta(2,2) density functions

Figure 3.4

how far he can deviate from this "equal probability rule", or what the effects of deviation might be.

Although the possibility of obtaining sensitivity information using Latin hypercube sampling is attractive, more research would be required to establish the method's practical usefulness.

A typical mathematical model may involve simulation of dozens of physical, chemical and cultural systems and mechanisms. Many of these phenomena are only imperfectly understood, and the parameters that specify them completely are not known with certainty in most cases. This implies that the confidence bands on model outputs will be very wide indeed. These confidence bands can be narrowed by improving our understanding of the processes that are being modelled, by improving the model's formulation of the processes, and by perfecting our estimates of the process parameters. It is natural to make use of the same mathematical model to try to determine ways in which our efforts to achieve these improvements can be expended most efficiently.

There are at least two sorts of questions that may be asked concerning how output prediction reliability is influenced by modelling reliability:

- At our present state of knowledge, what is the rate of change of prediction reliability with respect to the rate of improvement of our state of knowledge?

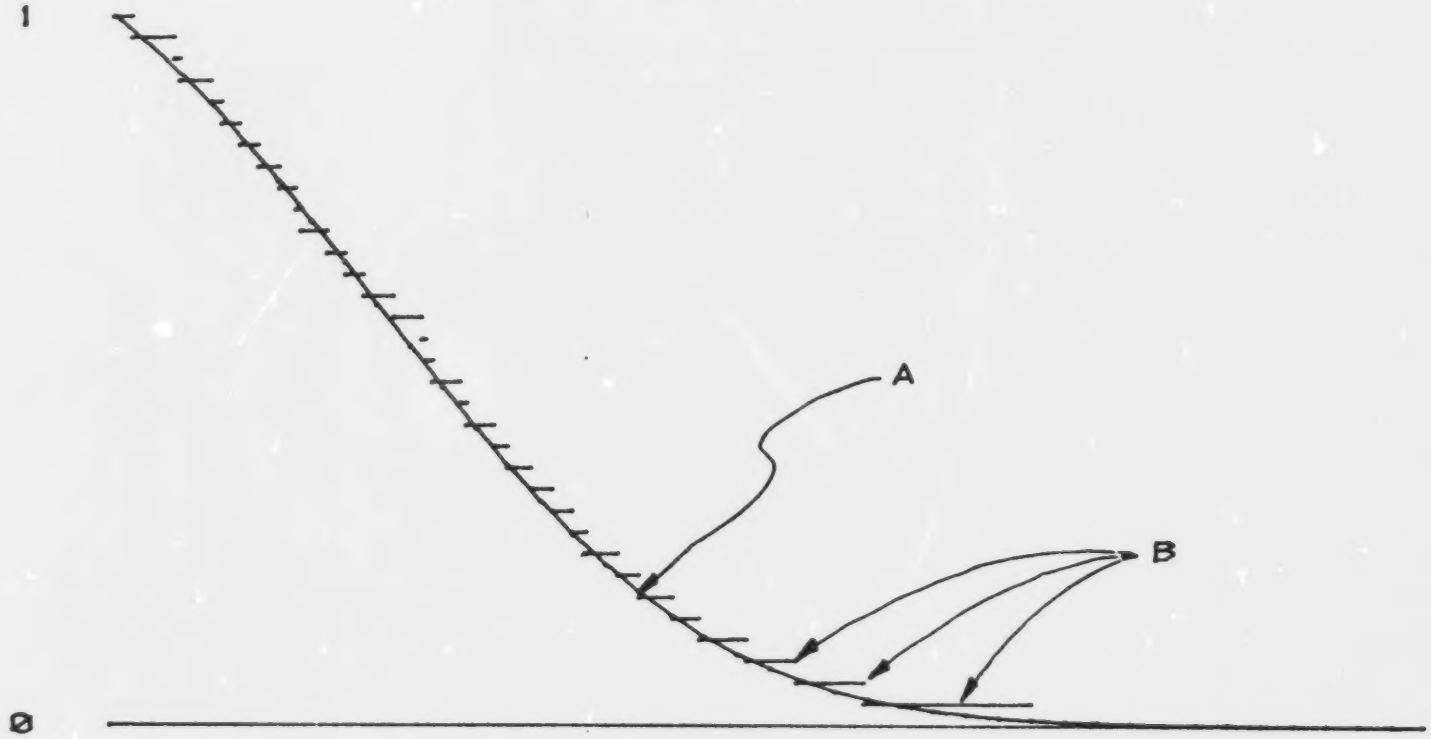
- How much prediction uncertainty is accounted for by each modelling uncertainty or collection of uncertainties?

Clearly, if prediction reliability is insufficient, then greatest effort should be applied to the study of those aspects of the system which apparently manifest the greatest reliability improvement rate of change, if there is a reasonable expectation that success will accompany such efforts. However, if the greatest proportion of prediction uncertainty is assignable to a knowledge of factors which cannot be significantly improved, then perhaps this suggests that other modelling approaches ought to be attempted.

There is an essential difficulty in this second approach: it will not be possible to partition the total variance on any one model output into a collection of variances associated with individual data and scenario uncertainties. Because of statistical nonorthogonality of the members of these sets of uncertainties, the variances associated with some of them will be inflated by contributions of variance from others. In other words, for realistically complicated models, it will be impossible to uniquely assign all components of the variance on any given model prediction to a subset of data and scenario uncertainties. However, even such imperfect assignments would be of use in making decisions concerning research directions.

Latin hypercube, and possibly full factorial, sampling can be used for assessment of rate of change of prediction reliability, but so can the adjoint method (if prediction reliability measured by variance is deemed acceptable). Latin hypercube sampling would appear to be of little use in assessing the total contribution of individual input uncertainties, for example, to prediction uncertainty but the adjoint method can likely be applied to this class of tasks.

In Section 5.0, we present some preliminary results involving questions of rate of change of prediction reliability with data reliability.



Typical survival function (A)

Sample empirical survival function (B)

Figure 3.2a

4.0 QUALITY OF CURRENT KNOWLEDGE RELATING TO MODELS OF ECOSYSTEMS SUBJECT TO DISCHARGES FROM URANIUM-MILL TAILINGS

4.1 Introduction

The quality of predictions made using a mathematical model depend, in part, on the quality of information relating to the model's parameters.

Three of IEC BEAK's experts[1] in the area of uranium tailings, were invited to write collections of brief prose descriptions of our knowledge of the parameters in the IEC BEAK mathematical model. These descriptions outline gaps and deficiencies in knowledge; they also sketch what work will be required to improve knowledge sufficiently to allow modelling for regulatory purposes.

The experts were also asked to formulate knowledge of the parameters in terms of probability density functions.

4.2 Assessing Expert Opinion in Terms of Probability Density Functions

The IEC BEAK mathematical model is given in terms of parameters about whose values we are uncertain. Expert knowledge of, and opinions about, these values are to be assessed using probability density functions. That is, experts are to be asked to render their knowledge and opinion as statements of probability. However, as Phillips (70) expresses it:

> "It is well to remember ... that a probability is a number between zero and one inclusive. I find it difficult to believe that my feelings of uncertainty register within me as numbers, to be detected, and .... reported honestly. My feels of uncertainties are feelings, not numbers, so they cannot be probabilities."

1 Dr. D.L. Lush, Principal, Mr. M.F. Holloran, Associate, Dr. W. Snodgrass: authors/co-authors of numerous reports in this area of research

The emphasis in this study was on studying methods for relating prediction reliability to various aspects of modelling reliability. However, we require a method (or methods) for turning "feelings" into "probabilities" — or probability density functions. Although methods which might suit our present purposes have been explored by statisticians and psychologists, we do not require results of the accuracy that these methods may achieve, and so the added expense of employing them would not have been justified for the present project. (For discussions concerning methods for elliciting expert opinion in terms of density functions see de Finetti (70), Hogarth (77), Jaynes (58), Kadane et al. (80), Schaeffer and Borcherding (73), Schmitt (69), von Holstein (70, 71), Tribus (62), Winkler (67a/b/c, 80), Winkler and Murphy (68)).

For our preliminary investigation of opinion about various parameters we asked experts to report their opinions about each parameter's likely values directly in terms of a beta probability density function.[1,2] This means that we asked them to specify the beta density function domains (ranges of possible parameter values), and to choose beta density parameter values, in such a way that the shapes of the beta density so chosen would approximate to their "mental pictures" of the densities that characterize their opinion. The experts were to "fold in" all sources of uncertainty to form each density function.

There were no theoretical grounds for our choice of beta densities.[3] We chose them because:

1 There are empirical grounds for serious doubt that experts can do this (Winkler (67a)).

2 See Appendix 2 for a definition of the beta density. See Hastings (75) or any other manual of probability distributions for more information.

3 The description of a model should include a definition of the region of the parameter space for which the model is valid (and for which it has been explicitly validated). Most projections of such regions onto subspaces spanned by subsets of the parameters are likely to be bounded sets. This implies that sampling for parameters from distributions over infinite supports is usually unacceptable. Use of a family of distributions with finite support, like the beta, is usually indicated. In this study, however, no attempt has been made to restrict the sample spaces on parameters to what might be considered feasible sets.

- beta densities assume a wide variety of shapes for various combinations of parameter values
- it is relatively easy to explain how various typical parameter value combinations give rise to typical density shapes.

Some fundamental concepts having to do with modelling parameter value predictability with density functions were reviewed with the experts, as an exercise to prepare them for their tasks. We discussed interpretations of various measures of location and dispersion, etc, and how these considerations related to the task which was at hand. We also provided an illustrated, verbal introduction to the properties of the beta density, and invited the experts to experiment with various parameter value combinations using a microcomputer programmed to display the resulting density functions. Finally, the experts were asked to draw our attention to cases of parameters whose reliability could not be adequately modelled using beta distributions.

The following points should be made:

- The experts were not asked to review published results concerning the parameters, or to consult with colleagues. These activities would be involved in any careful programme of expert opinion ellicitation, as would meticulous structuring and control of such a programme.

- No formal procedures were applied for reconciling opinions. It must be said, though, that scientifically derived methods for combining and reconciling opinions from a number of sources are not well developed.

5.0 APPLICATION OF A METHOD FOR PROBABILISTIC CALCULATION OF DOSE TO AN EXISTING MODEL

As mentioned earlier, a computer program (the IEC BEAK model) exists which embodies a mathematical model of an aquatic ecosystem which includes a uranium tailings pond. The program generates output which includes tables with the format shown in Figure 5.1.

This table presents dose levels for various populations in the ecosystem as a function of time. The "critical group" is that population of human individuals who live in the vicinity of the lake to which tailings seepage flows. This lake is located in a natural chain of waters which enter the Great Lakes. Thus, members of the critical group would appear to be the individuals subject to the greatest exposure to radiation in the ecosystem.

Critical group members under one year of age typically exhibit feeding habits and other behaviour which differ from those of group members over one year of age. For this reason, the exposures of the two age groups were modelled separately. Results typified by those in Figure 5.1 indicate that the older age group members receive higher maximum doses. In order to simplify exposition we choose to redefine "critical group" to mean "critical group excluding members less than one year of age" for the remainder of this section.

One of our tasks in this project was to modify the existing program so that it could "compute the frequency distribution of the maximum dose to typical members of the critical group, the doses themselves having been calculated as a function of time." (Coady(83))

The existing program was modified so that, instead of making calculations only for a single vector of input parameters, it repeated its calculations for a series of input parameter vectors chosen by a Latin hypercube sampling scheme of the kind described in Section 2.1.2 of this report. These modifications also determine the maximum dose over time for each of these input parameter vectors and collect the predicted maximum doses (model output) as an empirical cumulative frequency distribution. This output is hereafter called an empirical cumulative distribution function.

TIME (YEARS)	CRITICAL GROUP DOSE RATE (REM/YR) UNDER 1	"ADULT"	SERPENT RIVER CUMULATIVE (REM)	GREAT LAKES CUMULATIVE (REM)	NORTH ATLANTIC CUMULATIVE (REM)
25.0	.103	.209	.110E+04	.189E+04	812.
50.0	.911E-01	.187	.219E+04	.466E+04	.323E+04
75.0	.849E-01	.174	.318E+04	.744E+04	.669E+04
100.	.826E-01	.170	.414E+04	.102E+05	.107E+05
125.	.824E-01	.169	.508E+04	.129E+05	.150E+05
150.	.828E-01	.170	.603E+04	.157E+05	.194E+05
175.	.276	.251	.938E+04	.231E+05	.259E+05
200.	.216	.439	.126E+05	.336E+05	.381E+05
225.	.165	.333	.146E+05	.401E+05	.510E+05
250.	.173	.352	.165E+05	.455E+05	.621E+05
275.	.163	.328	.184E+05	.507E+05	.719E+05
300.	.177	.360	.203E+05	.559E+05	.810E+05
325.	.184	.378	.225E+05	.612E+05	.901E+05
350.	.191	.381	.245E+05	.667E+05	.992E+05
375.	.140	.260	.266E+05	.720E+05	.108E+06
400.	.114E-01	.233E-01	.270E+05	.747E+05	.116E+06
425.	.841E-03	.173E-02	.271E+05	.752E+05	.121E+06
450.	.636E-04	.130E-03	.271E+05	.753E+05	.123E+06
475.	.527E-05	.105E-04	.271E+05	.753E+05	.125E+06
500.	.576E-06	.105E-05	.271E+05	.753E+05	.125E+06
525.	.669E-07	.162E-06	.271E+05	.753E+05	.125E+06
550.	.237E-07	.376E-07	.271E+05	.753E+05	.126E+06
575.	.659E-08	.105E-07	.271E+05	.753E+05	.126E+06
600.	.197E-08	.308E-08	.271E+05	.753E+05	.126E+06
625.	.585E-09	.917E-09	.271E+05	.753E+05	.126E+06
650.	.174E-09	.273E-09	.271E+05	.753E+05	.126E+06
675.	.515E-10	.813E-10	.271E+05	.753E+05	.127E+06
700.	.152E-10	.242E-10	.271E+05	.753E+05	.127E+06
725.	.451E-11	.717E-11	.271E+05	.753E+05	.127E+06
750.	.133E-11	.212E-11	.271E+05	.753E+05	.127E+06
775.	.392E-12	.628E-12	.271E+05	.753E+05	.127E+06
800.	.116E-12	.186E-12	.271E+05	.753E+05	.127E+06
825.	.340E-13	.547E-13	.271E+05	.753E+05	.128E+06
850.	.999E-14	.161E-13	.271E+05	.753E+05	.128E+06
875.	.294E-14	.473E-14	.271E+05	.753E+05	.128E+06
900.	.861E-15	.140E-14	.271E+05	.753E+05	.128E+06
925.	.253E-15	.410E-15	.271E+05	.753E+05	.128E+06
950.	.740E-16	.120E-15	.271E+05	.753E+05	.128E+06
975.	.217E-16	.353E-16	.271E+05	.753E+05	.129E+06
.100E+04	.634E-17	.103E-16	.271E+05	.753E+05	.129E+06

FIGURE 5.1

The table that the computer program supplies as a representation of the empirical cumulative distribution function is not particularly well suited to human interpretation or appreciation. Furthermore, we will see that an investigation of the sensitivity of this output characterization to input distribution variation involves repeated calculation of frequency distribution estimates. This implied that small adjunct computer programs should be written to re-process output from the computer modelling program. We constructed two programs to provide displays of an empirical distribution function and an empirical frequency polygon. These programs are described briefly in Appendix 1.

The purpose of this study was to investigate methods for quantifying reliability, rather than to actually complete a reliability study of an existing model. Thus it was agreed that only a relatively small number of Monte Carlo samples would be required to demonstrate a method. Unfortunately, even the most refined statistical methods for probability density function estimation require quite large samples to produce results which are at all acceptable. For this reason, applications of the program for graphing empirical density functions do not appear in this report (other than in the form of an example).

In the next section, we review data and scenario reliabilities for the IEC BEAK model; we also present a table of the parameters of the beta distributions which were used to model data uncertainty. In Section 5.2 we present the results of our Monte Carlo study.

5.1 Expert Opinion Concerning Data and Scenario Reliability for the IEC BEAK Computer Model

The notes in this section of the report are a compilation of the opinions of the three experts identified earlier, and of the primary author of the report, W. Bell. Opinions were rendered concerning various aspects of model, scenario and data reliability, as discussed in Section 1.2, in terms of the four collections of parameters in the model, namely loading parameters, lake parameters, food chain parameters and dose conversion factors. The loading parameters describe the tailings pile, the lake parameters describe the ecosystem of each lake through which tailings leachate flows, and the dose conversion factors describe how human exposure is related to environmental parameters. "Food chain parameters" is something of a misnomer in that these parameters describe exposures via swimming, sunning, etc., as well as dietary items.

Scenario Reliability

The computer model does not include representations of massive changes such as glaciation cycles, meteoric impacts, or the effects of changes in cultural influences. Examples of changes not included are as follows.

- **Loading Parameters:** Changes to ore-processing technology, or metallurgical developments in the future could radically change the chemical and physical behaviours of uranium mill tailings. Improved methods of tailings treatment in future might reduce the amounts of radionuclide contaminants that reach the ecosystem. Long-term cycles in atmospheric precipitation levels could alter the behaviour of tailings.

- **Lake Parameters:** Long-term changes in atmospheric precipitation levels could alter lake volumes and residence times. Man may change the topology of the system of waters in building systems of dams or canals.

- **Food Chain Parameters:** The lands around the bodies of water in the model might be turned to different uses in future. For example, some of the land might see increased recreational use, or some of it may become valuable for housing or industrial uses. Some of the species which are included in the diets of peoples living around the lakes may be lost. There is likely to be change in the cultural composition of the peoples that live around the lakes simulated in the model.

Model Reliability

One of the most salient features of the IEC BEAK model is that it was designed to simulate hypothetical, but typical, ecosystems over long periods of time. To a degree, this reduces its reliability over shorter time spans since few short-term mechanisms are included in the model. There is not at present an explicit description of the parameter space for which the model is valid.

It is possible for unfeasible combinations of parameter values to arise during the operation of the model. A first step toward eliminating such occurrences would be to

program more exhaustive diagnostics into the computer code. When problems are diagnosed other parts of the code would be enhanced to obviate them.

- **Loading Parameters:** Tailings adsorption coefficients are not known with certainty but accommodation of variability in this model subsystem involves structural change to the program.

- **Lake Parameters:** The computer model assumes zero background radionuclide concentrations. Typical values could range from less than 1 to 10 pCi/litre depending upon the isotope. Variability in effective distribution coefficients might be expected owing to uncertainty in lake chemistry; additional structure would have to be included in the model to account for these variabilities. The model does not allow for flow balance considerations, but lake surface areas and volumes depend on the pattern, sizes and relative sizes of flows within the system. The model assumes that the area of the sediment in a lake is the same as the surface area of the lake; in fact, sediment area is likely to be somewhat larger than surface area in most cases. The model does not take into account the fact that there must be seasonal and spatial variations in mixing in the lakes that it models.

- **Food Chain Parameters:** The model does not account for population size changes. Neither does it consider the effects on other populations of shipments of food products from the area.

Data Reliability

- **Loading Parameters:** The tailings represented in the model is a hypothetical one only; actual tailings can be expected to differ in various ways. Component chemical concentrations: It is necessary to specify a wide range of initial chemical conditions within the tailings mass. These chemical concentrations are derived from values measured both within the mill and within the tailings themselves. Within the tailings, concentrations of all elements are well above analytical detection limits and the error associated with their derivations, both analytical and calculated from mass balance considerations, may be expected to be relatively small. Infiltration rate through tailings depends on tailings surface

area, evaporation rate and rate of rainfall. It is thus site-specific, and also subject to temporal variation. The gypsum solubility coefficient depends on the degree and quality of gypsum crystallization in the mill product, and whether the crystals become coated or encapsulated. These depend upon such variables as mill operating conditions, the nature of the ore, and geochemical processes in the tailings. The upper limits on porewater thorium are theoretical guesses. These values are used to estimate leaching rates only during the period that the tailings porewater remains basic. Tailings porosity is related to the tailings solids and to dissolution characteristics. Values are thus site-specific, and are subject to spatial and possible temporal variation. There is no precise theory for the value used for the time to degrade one-fourth of the jarosite load. However, there are some data which might be used to establish a value. Masses of nuclides originally present in ore were taken from ore assay information. Thus their reliabilities depend on the qualities of these assays, and the characteristics of the ore body from which the tailings were taken. Hydraulic flow through tailings is site-specific. It depends on size of drainage basin, whether watershed flows are routed through the tailings or around it and the premeability of the tailings. The value chosen was intended to be typical.

- Lake Parameters: Variations among lakes are provided for in the computer model, but there are often insufficient data to be able to provide different values for different lakes. The uncertainty in the measurement of a lake volume relates to how complex the bathymetry of the basin is and the level of detail that is put into the bathymetric profiles. In a general way, if we assume the bathymetric complexity to be similar, we may expect a higher percentage error associated with a small surface area water body than a large one. The range for the lakes modelled in this report may range from 1-5% for the large water bodies (such as the Great Lakes) to close to 30% for some of the small system lakes where volume is estimated based upon surface area and estimated mean depth. Lake mean depths: Their reliabilities are influenced by the same factors that influence lake volume reliabilities. Lake residence times are calculated from lake volumes and flow; hence, residence time reliability depends on those for volume and flow. The distribution coefficient is a measure of the ratio of radionuclide concentration in the aqueous phase and the concentration in the solid phase. The uncertainty associated with this value will depend upon the

uncertainties in the component parts of the ratio. If the concentration in the solute and solid are both low, the overall ratio uncertainty could be very high. If the component concentrations are both high the overall uncertainty would be lower. To estimate the uncertainty for any distribution coefficient it is necessary to review the original data used to derive it and not simply the published number. Lake discharge measurements can be based upon actual measurements of discharge or watershed yields. The uncertainty associated with the value used will depend upon the period of record over which the data were collected and the accuracy with which the data were collected. Since the data used in the particular model are based upon information contained in an Ontario Ministry of the Environment model, we have assumed high reliability. The settling velocity of the mass of particles in a lake on a long-term average depends upon the particle size distribution of the particulates in the water column. The smaller the particles, the more slowly they settle. The value chosen of 25 m/year is an average surface area normalized settling velocity. The distribution on velocities may be expected to be skewed toward lower velocities to reflect the presence of both biological particulates and aggregation of colloids. In some lakes, there are particulates that do not settle. Sedimentation rate is a lake-specific property which depends on whether the drainage basin is cleared or not. There are insufficient data to provide distinct values for each lake. Determinations of diffusion rates within sediments are based upon empirical measurements usually well above the detection limits of the elements concerned. Accordingly, it is assumed that the distribution will have a fairly narrow spread for determinations made in the laboratory. Lake sediment area is assumed to be related to surface area and depth. The sediment surface area uncertainty is expected to be directly related to the surface area uncertainty which, in turn, is related to the complexity of the basin (that is, number of shoals, islands, etc.), and the level of mapping detail provided. The sediment-to-beach transfer gives the percentage of beach material which consists of "lake sediments", that is, those with adsorbed radionuclides, and the percentage which consists of terrestrially derived alluvium. Sediment porosity will depend upon the composition of the material (that is, sand, silt, clay or organic material), and its degree of consolidation. For this term, we assume a distribution which is generic for the headwater lakes and measured for the Great Lakes. The same rationale is proposed for sediment bulk density. Sediment depth: The depths of

the aerobic and anaerobic layers in sediment in which diffusional processes and chemical reactions play an important role with respect to resultant overlying water concentration vary from lake to lake and are dependent upon a variety of environmental factors.

- **Food Chain Parameters:** The figures included in the computer model are for hypothetical populations only. No allowances for individual differences or cultural factors are made. In any case, though, there is little information available which quantifies these behaviours. The bioconcentration factors have many of the same uncertainties associated with them as the distribution coefficients for sediments and suspended particulates. The exact nature and magnitude of uncertainty will depend upon how the factor was determined rather than upon its absolute value. The numbers of persons participating in an activity such as swimming or sunbathing are assumed to be constant proportions of populations. It is, however, likely that the smaller the lake surface area the higher the percentage of the nearby population group that participate, at least as far as the ocean. Accordingly, the uncertainty would be largest with the large area great lakes and smaller with the smaller headwater lakes. There seems to be few actual data on how much time people spend at the various activities. Internal exposure to radiation results from the ingestion of foodstuffs and water. Dose conversion factors (discussed below) are used to convert these exposures to radiation doses. Any given individual may consume more or less than this amount. There are definite limits for maximum and minimum consumption. External exposure: Individuals may be exposed to ionizing radiation from direct exposure to gamma radiation from ground, water or air. The total amount of time spent either walking on contaminated ground (sediments in our model) or immersed (swimming and bathing) probably has a distribution associated with it with a large dispersion. Exposure can also result from breathing of contaminated air. Again, the amount of air inhaled probably has much the same shape of probabililty distribution as water consumed.

- **Dose Conversion Parameters:** The accepted values for these parameters have changed even over just the past few years as we develop better analytical techniques and come to a better understanding of the systems involved. The dose conversion factors employed for each of the radionuclides are based upon a set of relatively complex metabolic models. Some radionuclides have a much broader experimental model data base associated with them than others.

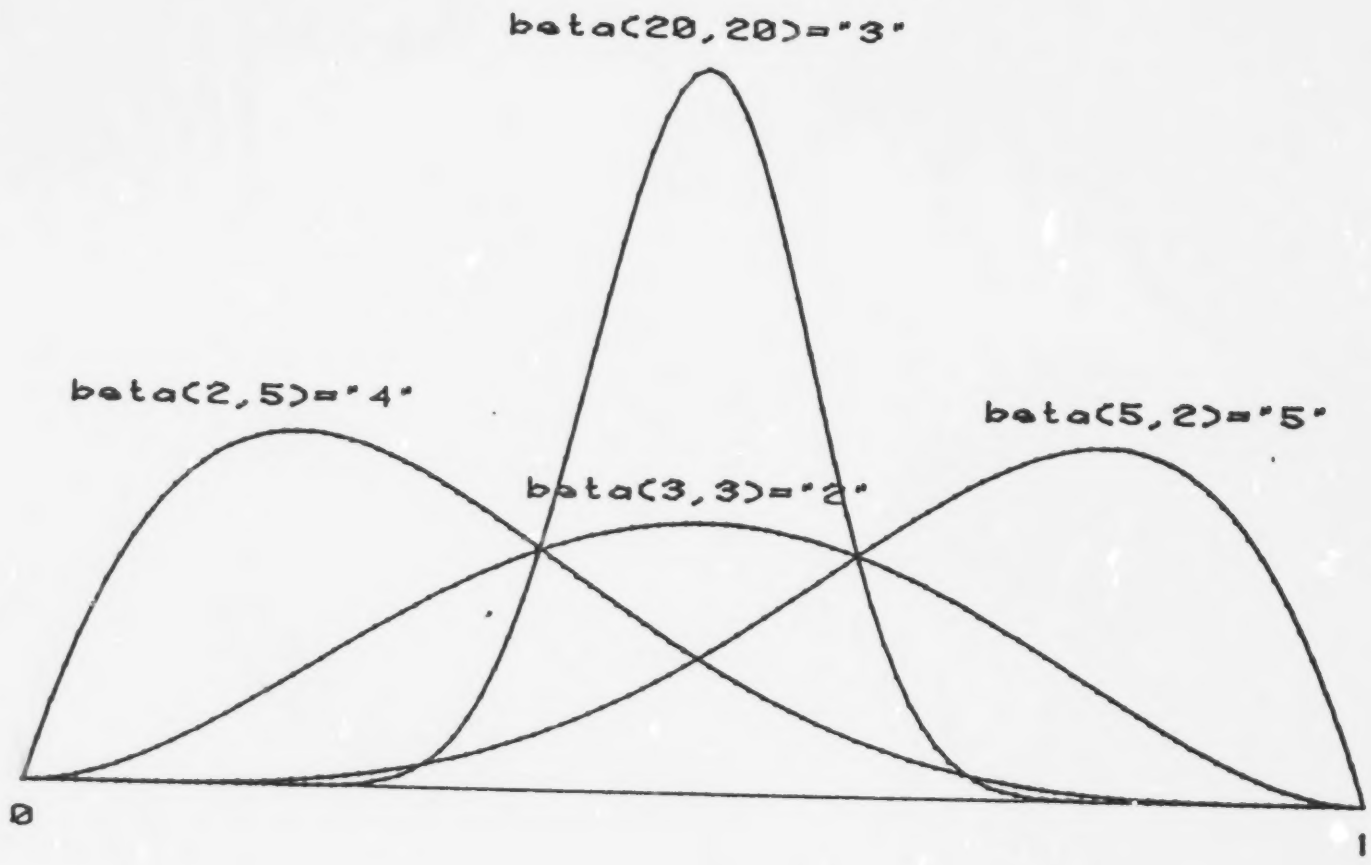
The experts were also asked to assess knowledge of the model's parameters in numerical terms. The experts chose from among five beta distributions to give their opinions about the shapes of the distributions representing their opinions. We have arbitrarily labelled these distributions as "Shape 1" through "Shape 5". The following table shows the correspondence between "shape" and beta density parameters.

Table 5.1: Correspondence between "shape" and beta density parameters

Shape	beta density parameters
1	1,1 (uniform distribution)
2	3,3
3	20,20
4	2,5
5	5,2

These densities are displayed in Figure 5.3.

Table 5.2 presents the combined expert opinions in numerical terms.



Beta density functions

Figure 5.3

TABLE 5.2: EXPERT ASSESSMENT OF MODEL'S PARAMETERS IN NUMERICAL TERMS

Parameter/Parameter Class	Category	Range	Dist'n Shape
Loading Parameters			
Total mass of radionuclides	U-238	.95, 1.05	1
	U-234	.95, 1.05	1
	Th-230	0, 2	2
	Ra-226	0, 2	2
	Pb-210	0, 2	2
	Po-210	0, 2	2
	Th-232	0, 2	2
	Ra-228	0, 2	2
	Th-228	0, 2	2
Fraction in gypsum	U-238	.01, 2	5
	U-234	.01, 2	5
	Th-230	.9, 1.05	2
	Ra-226	.9, 1.05	2
	Pb-210	.9, 1.05	2
	Po-210	.9. 1.05	2
	Th-232	.9, 1.05	2
	Ra-228	.9, 1.05	2
	Th-228	.9, 1.05	2
Fraction in jarosite		.9, 1.05	2
Total mass originally present in ore		.8, 1.2	3
Porosity		.75, 1.3	2
Adsorption coefficients		.5, 1000	3
Pyrite oxidation rate		.1, 100	4
Tailings pyrite mass percent		.4, 1.2	2

TABLE 5.2: EXPERT ASSESSMENT OF MODEL'S PARAMETERS IN NUMERICAL TERMS (Cont'd)

Parameter/Parameter Class	Category	Range	Dist'n Shape
Hydraulic flow rate	Base case	.01, 5	2
Infiltration rate		.5, 1.5	2
Porewater concentrations	Uranium	.25, 2	2
	Thorium	.01, 50	2
	Radium	.1, 5	2
	Lead	.1, 10	2
	Polonium	.1, 10	2
Gypsum concentration		.25, 1.2	2
Calcium concentrtion		.25, 1.2	2
Gypsum		.1, 10	3
Porewater constraint on thorium		.1, 10	1
Mean porous zone depth		.5, 1.5	2
Mean tailings depth		.6, 2	2
Tailings mass		.5, 10	1
Time to degrade .25 of jarosite		.04, 40	1
<u>Lake Parameters</u>			
Suspended sediments	.133 values	.25, 4	1
	.667 values	.25, 2	1
	1.33 values	.1, 3	1
	.33E-2 values	.5, 2	1
Lake discharge	headwater lake system	.1, 5	5
	Others	.5, 1.5	2
Nuclide decay ratio		.95, 1.05	3
Lake volume		.9. 1.1	3
Particle settling velocity		.004, 8	5

TABLE 5.2: EXPERT ASSESSMENT OF MODEL'S PARAMETERS IN NUMERICAL TERMS (Cont'd)

Parameter/Parameter Class	Category	Range	Dist'n Shape
Diffusion coefficient		.25, 5	2
Lake surface area		.95, 1.05	3
Sedimentation rate		.05, 2.5	4
Adsorption coefficients		.9, 10	4
Mean aerobic sediment depth	Shield Lakes	.25, 2.5	4
	Great Lakes	.1, 2	5
Mean anaerobic sediment depth		.5, 100	4
Sediment porosity		.8, 1.2	2
Sediment length scales		.1, 10	2
Particle adsorption coefficient		.001, 100	2
Dose Conversion Factors			
All		0, 2	1
Food Chain Parameters			
Transfer coefficients		.1, 10	2
Annual consumption rate	Water: critical group	.1, 2	2
	Water: others	.75, 1.25	2
	Fish: non North Atlantic	.25, 1.25	2
	Fish: North Atlantic	.5, 1.5	2
	Moose: headwater river and lake System	0, 2	2
	Wild rice	0, 2	2

TABLE 5.2: EXPERT ASSESSMENT OF MODEL'S PARAMETERS IN NUMERICAL TERMS (Cont'd)

Parameter/Parameter Class	Category	Dist'n Range	Shape
Harvest rate	Water	.5, 1.5	3
	Fish: headwater lakes	0, 1.25	3
	Fish: Great Lakes	.25, 1.25	3
	Fish: North Atlantic	.25, 1.25	3
	Moose, Shellfish	0, 2	3
# people swimming/bathing in waters		.5, 1.5	1
# people sunning on water's shores		0, 2	1

For each parameter or collection of parameters the "range" in the table is a pair of numbers which give the interval from which the parameter were to be sampled in the Monte Carlo run. For example, the pair (.2,3) would indicate that the sampling limits for the parameter are .2 times the experts' "best" value for the parameter, and 3 times this best value respectively. The best values are those which were already incorporated in the software.

5.2 Results of Monte Carlo Study

In this section, an application of Latin hypercube sampling to model parameters is presented. This application takes the form of a preliminary examination of the sensitivity of model outputs to perturbations of parameter probability distributions.

The unreliability of many of the parameter values is reflected in the wide ranges of possible values for them given in Table 5.2. The computer model proved insufficiently robust to cope with extreme values for some of the parameters[1], and consequently it was decided that sensitivity investigations should begin with a study of the model's behaviour under narrower deviations from the "best parameter values".

Accordingly, for the first trial, all 1,100 or so model parameters were chosen from distributions with the beta(3,3) shape, with ranges of from .95 times best parameter value to 1.05 times this value. The model outputs for these distributions displayed insufficient variation for a demonstration of sensitivity investigations using the Latin hypercube. In fact, of N=10 maximum critical group dose rate outputs, two were equal, at .391 rem/a, and the remaining seven were equal, at .447 rem/a.

1 A thorough investigation of model program design and coding would have been beyond the scope of the work reported here. The deficiencies in robustness may owe to numerical algorithm instability, or to coding which does not allow for all eventualities. Needless to say, these problems will have to be eliminated before more extensive sensitivity studies can be undertaken.

In the next trial, the variabilities of all model parameters were increased. The beta (3,3) shape was retained, but the ranges used were from .9 times best parameter value to 1.1 times this value. Figure 5.4 is a reproduction of part of the computer output for this trial. The APL program EDF_DRAW, described in Appendix 1, was used to produce the display of the estimate $\hat{S}_{lhs}(.)$ shown in Figure 5.5. (For reasons mentioned in Appendix 1 we have not provided an estimate of the probability density function on maximum dose.)

Since N = 10 Latin hypercube samples were to be drawn, each model parameter's range was divided into 10 subranges each with probability .1, that is, equally probable ones, under the beta (3,3) distribution.

As an illustration we now describe a preliminary review of the sensitivity of the IEC BEAK model to changes in the distribution shapes on some of the model's parameters, from beta (3,3) to beta (2,2). If expert opinion about these parameters could be given as a beta (3,3) distribution then, since the beta (2,2) distribution has greater variance than the beta (3,3), this amounts to entertaining the possibility that knowledge of some parameters is not as good as is believed.

Figures 5.6 and 5.10 are displays of perturbed estimates of $\hat{S}_{lhs}(.)$, the arbitrarily-chosen variables for perturbation being lake volumes, lake surface areas, aerobic sediment depths, anaerobic sediment depths and lake depths respectively.

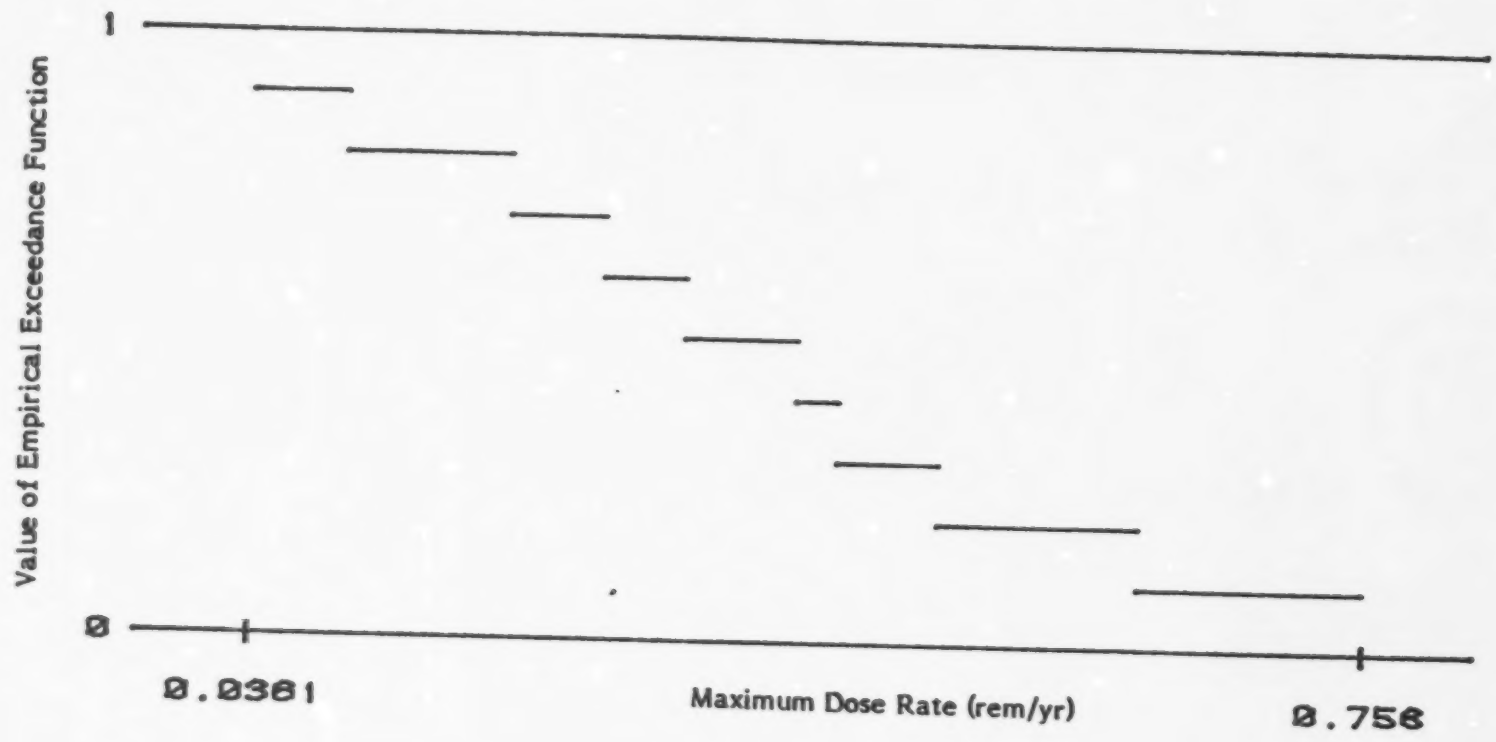
The most salient feature of the graphs is their lack of variability. Of course, the positions of the salti (or "jumps"), along the horizontal axis, do not vary over the graphs, in line with the definition of $\hat{S}_{lhs}(.)$ and perturbed versions of it. The only variations are in the heights of the salti: these variations are reflections of the differing contributions to probability owing to variations from equal probability of sampling from the distributions on the model parameters. Since numerous parameter distributions are perturbed at once the variations in probability tend to cancel one another out.

An apparent option is to perturb only a small number of distributions at one time, but, with this approach, the contribution of such a parameter must be virtually negligible. Clearly, displays of perturbed estimators cannot be expected to yield much information, in themselves, about a model's sensitivity to model parameter distribution perturbation. This is a considerable problem.

EMPIRICAL EXCEEDANCE FUNCTION
MAX DOSE TO CRITICAL GROUP OVER TIME

ABSCISSAE	ORDINATES
-INFINITY	1.00
.361E-01	.900
.965E-01	.800
.203	.700
.264	.600
.317	.500
.389	.400
.416	.300
.482	.200
.612	.100
.758	0.

FIGURE 5.4



Maximum Dose to Critical Group Over Time

Empirical Exceedance Function

FIGURE 5.5

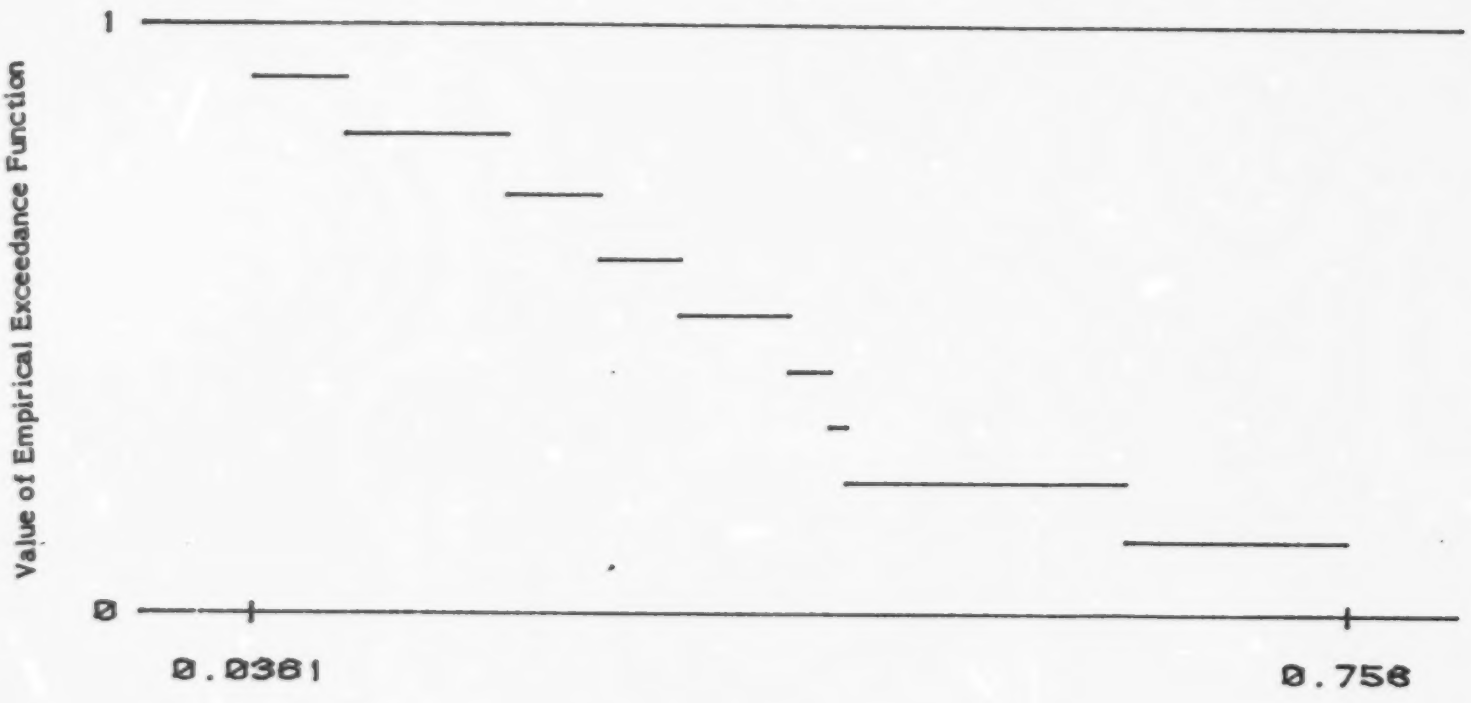


FIGURE 5.6: Perturbed estimates of estimated exceedance function for perturbation of lake volume

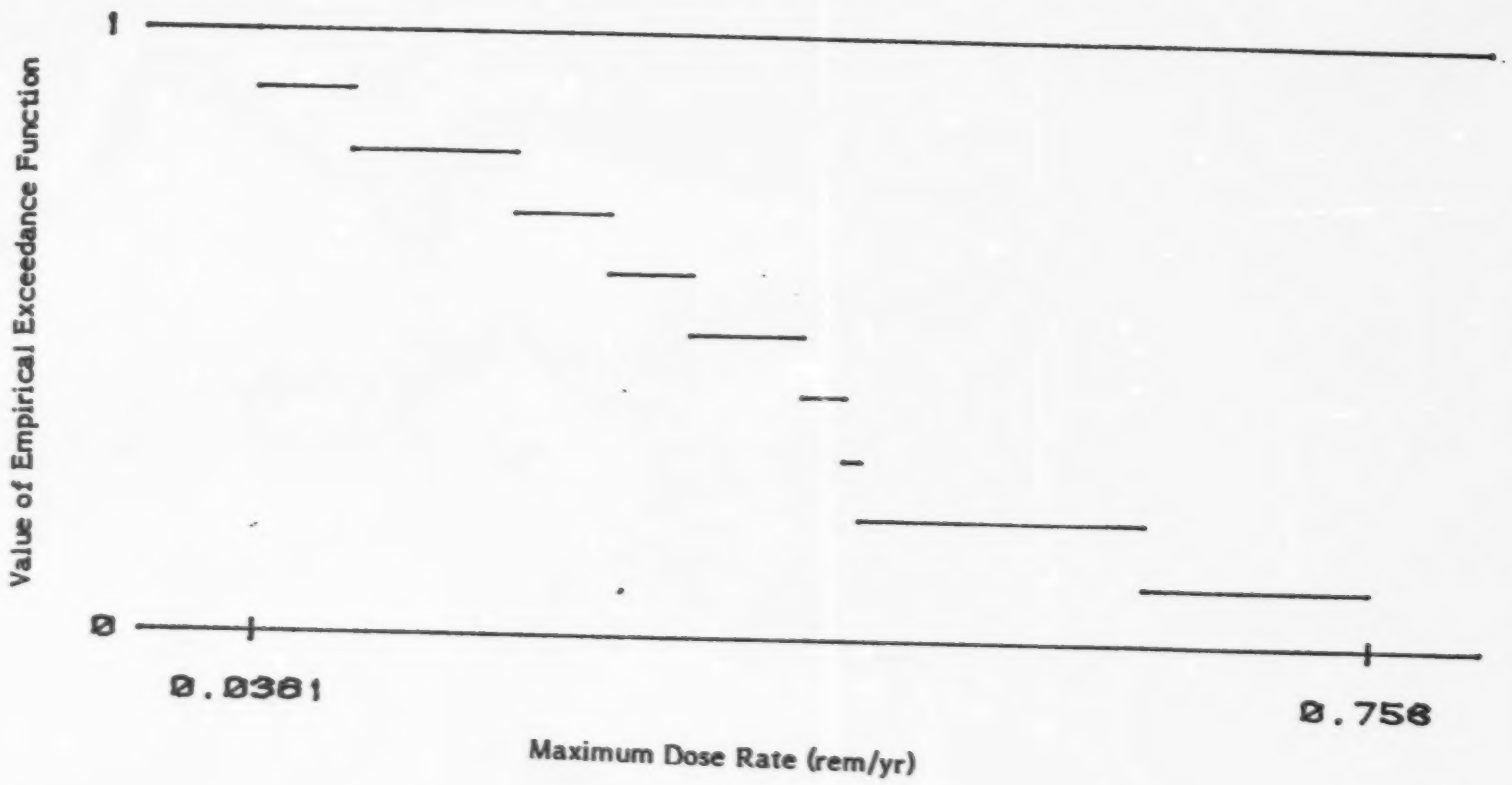


FIGURE 5.7: Perturbed estimate of estimated exceedance function for perturbation of lake surface areas

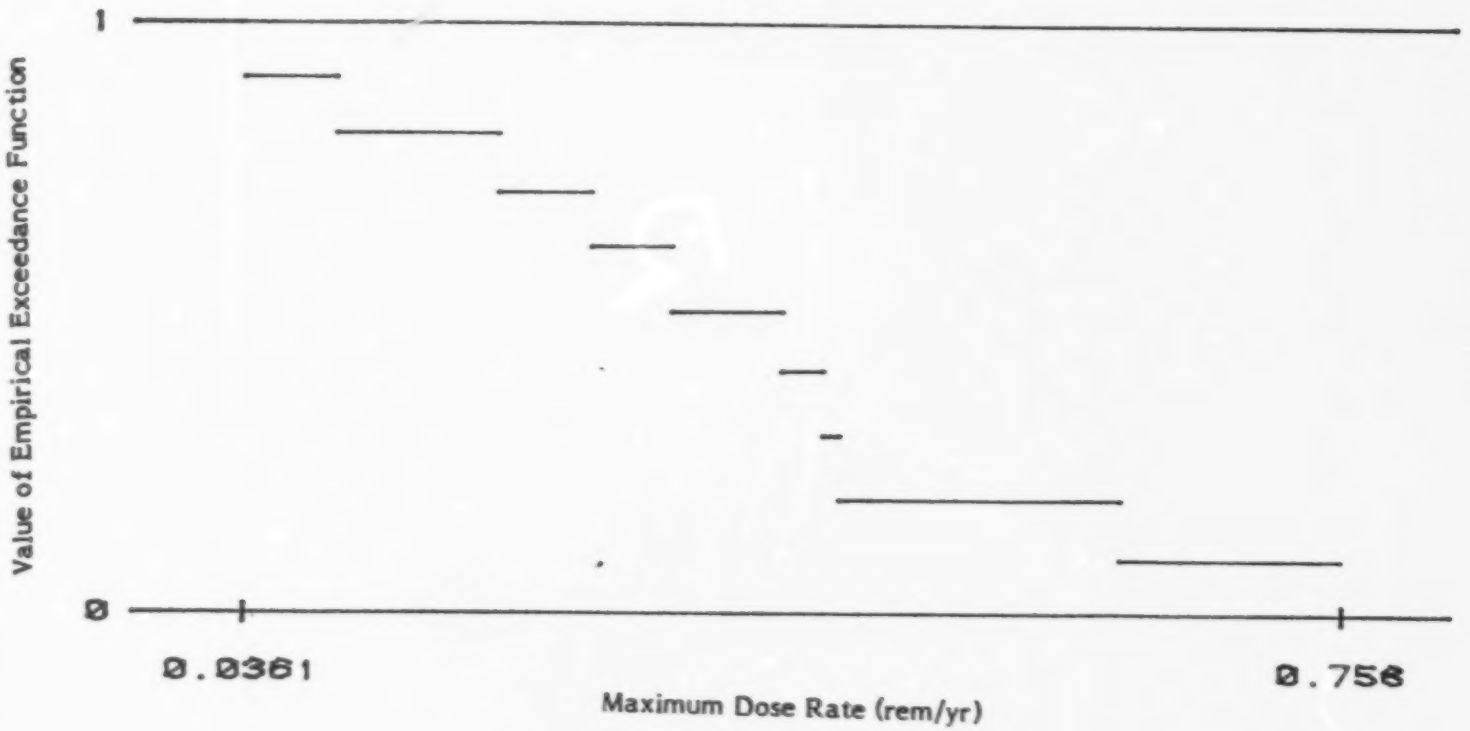


FIGURE 5.8: Perturbed estimate of estimated exceedance function for perturbation of aerobic sediment depth

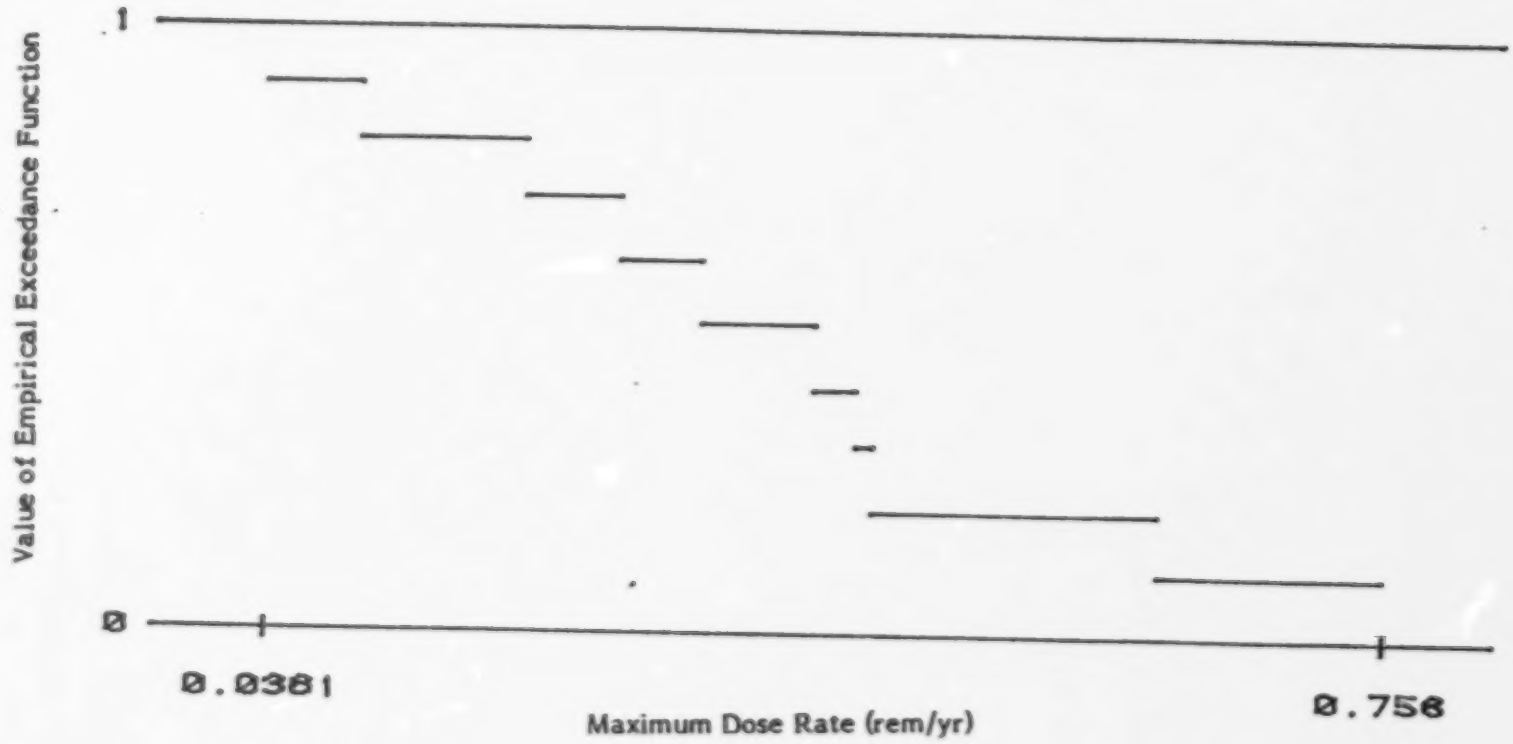


FIGURE 5.9: Perturbed estimate of estimated exceedance function for perturbation of anaerobic sediment depth

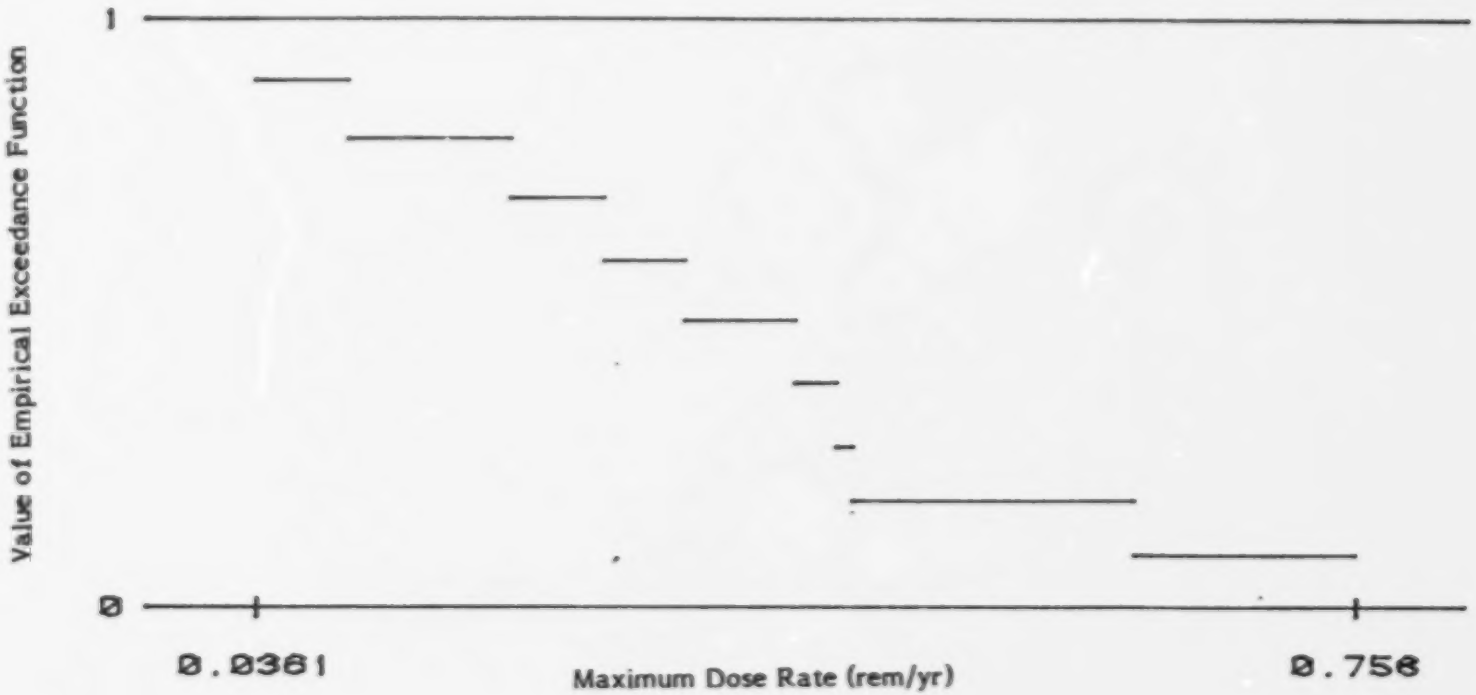


FIGURE 5.10: Perturbed estimate of estimated exceedance function for perturbation of lake depth

One approach is to analyse the results of perturbations statistically, but there are a number of difficulties and flaws in this approach. Firstly, the perturbed results can be treated as if they constitute a random sample but this is unacceptable on theoretical grounds inasmuch as the perturbed results are not actually obtained as a random sample. Secondly, the results can be pursued using pseudo-statistical methods but the problems inherent in this approach are combinatorically formidable where more than a few model parameters are involved.

We emphasize that these sensitivity results must not be relied upon, in view of the very small Monte Carlo sample size. Dependable results could only be obtained using, say, 1,000 samples, or more.

6.0 SOME CONCLUDING REMARKS

The predictions and summaries of a system's behaviour obtained using mathematical modelling are subject to a set of uncertainties which can be placed in three categories: data uncertainties, scenario uncertainties and model uncertainties. This study addressed some of the ways in which the effects of these uncertainties on model results can be assessed and quantified.

A realistic mathematical model of a complicated system typically involves hundreds of parameters and scores of differential and state equations and constraints. This usually means that knowledge of parameters and other features of the model is incomplete. Because such information is incomplete, and because it is usually only experts that understand model subsystems and their parameters properly, the only source of reliable, coherent information about the elements of a model is the collection of experts that are familiar with these elements. Expert opinion about the "true" values of model parameters can be expressed in terms of probability density functions. Thus, one way to assess the reliability of results obtained using mathematical modelling is in terms of the probability density functions rendered as the opinions of experts.

Accurate assessment of expert opinion is of fundamental importance in studying the reliability of a model's results because the experts' uncertainties about parameter values "propagate through" the model to evince themselves as uncertainties in the model's output. In other words, the model merely transforms uncertainty about input parameters into uncertainty about its outputs. In spite of the fundamental importance of accuracy in this area of accounting for uncertainty, development of rational, empirically-verified methods for assessing opinion concerning parameters of importance to radioactive waste disposal are in a rudimentary state of development.

It seems unlikely that an expert holds an opinion about any one parameter in isolation from his opinions concerning others. If opinions about parameters are gathered one-at-a-time, however, with no consideration being given to this characteristic of opinion patterns, then it is inevitable that some parameter combinations, considered impossible by experts, will be entertained during a model reliability study. This implies that other parameter-value combinations, which would be considered highly probable by experts, will not be given due consideration. Thus, in order to model expert opinion about model

parameters we must be prepared to ellicit their opinions in terms of multivariate density functions on the parameters. If methods for ellicitation of opinion in terms of univariate probability density functions are in a rudimentary state of development then methods for the multivariate analogue of this problem must be almost nonexistent.

Experts, and the organisations that gather information from them, are rarely impartial. Furthermore, it would be possible to falsify expressions of opinions in terms of density functions so that some "acceptable" density distributions on model outputs is obtained. For these reasons, regulators ought to prepare themselves with objective methods for gauging the authenticity and veracity of density functions purported to be expressions of unbiased expert opinion.

Although an expert may be confident that he has accurately rendered his opinion in terms of a probability density function, it is natural for a responsible observer to entertain questions concerning the "reliability" of such a density function. The observer will want to try to express this reliability mathematically so that he can determine the nature and degree of model output reliability upon the reliability of the expert's density function. Essential difficulties are, however, encountered in pursuing this line of thought. These difficulties are sketched in the following few paragraphs.

Suppose that we entertain the possibility that an expert provides opinions in terms of density functions that are one-element samples[1] from a population of such functions. We harbour the feeling that expert opinion can be subjected to some sort of probabilistic reliability analysis, and this paradigm certainly provides for a kind of random uncertainty in the expert's opinion. But the paradigm is flawed.

[1] It is conceivable that an expert could be induced to provide somewhat larger samples from his density function populations. Aside from the fact that it seems implausible that an expert could render a number of "best opinions," though, such a collection of density functions would still amount to a single expression of opinion about a parameter value because the information content of the collection could be given in terms of a single density function. Therefore, let us assume that an expert provides a one-element sample from his internal population of density functions.

Given a single density function from an expert one has virtually no information to use in doing a statistical reconstruction of the population. Valid statistical inference about the population of density functions from which the expert is assumed to have made his choice are therefore impossible.

One could, on the other hand, partially define the population from which the expert drew his opinion, and then use his density function to complete the definition using some estimation technique. The objection to this procedure is that there is no objective information upon which to base such a partial definition, and, anyway, the only kind of individual who could be called upon to provide such a definition would be another expert, and then one would have to begin again to consider the quality of his opinion.

Let us conclude by saying that, if an expert has accurately and honestly rendered his opinion in terms of a probability density function, then this is his opinion, and any uncertainty that the expert feels concerning the paradigm at issue must have been enfolded in the density function. If expert opinions cannot be subjected to probabilistic reliability analysis then more research should be conducted into applying and devising methods for reconciling and summarising the opinions of collections of experts with knowledge of uranium mill tailings management systems and the associated ecosystems.

Experts are commonly asked to describe their opinions concerning parameter values in terms of probability density functions. A mathematical model maps these density functions into distributions over the model's outputs. The distributions on parameters and outputs reflect the uncertainties with which the model parameters and outputs are known. Note, however, that since the distributions on the inputs cannot be considered to be probabilistically uncertain (for reasons just discussed), there can be no probabilistical uncertainty, ascribable to inputs, associated with the true output distributions.

Although the "true" output distributions are not subject to probabilistic variability they cannot be calculated with exactitude. This owes to the fact that realistic mathematical models are so complicated that the mathematical manipulations involved in deriving output distributions from input distributions are almost always intractable. A number of methods are in use, or have been proposed, for circumventing these problems of intractibility. In this report we introduced some variants of the Monte Carlo methods, and we mentioned the adjoint method briefly.

The Monte Carlo methods use probabilistic sampling and statistical estimation procedures to provide estimates of model output distributions. Consequently, the output distribution estimates include components of sampling variation even though the true output distributions do not. The magnitude of the sampling variation can be controlled by using Monte Carlo samples of sufficient size. However, just what sample size is sufficient for a particular, useful combination of sampling scheme and estimation procedure is usually unknown. Furthermore, almost all of the sampling schemes proposed for use in Monte Carlo experimentation, such as the Latin hypercube, are deficient in that they do not provide for sampling of the entire model parameter space. (Other sampling schemes are generally too expensive to use.) This means that parameter-value combinations of high consequence can easily escape adequate scrutiny, and, therefore, we do not believe that results obtained from the use of such fractional sampling plan Monte Carlo methods provide reliable results, in applications involving estimation of the output distributions associated with mathematical models of uranium mill tailings systems.

Again, there is a component of sampling variation in the output distribution estimates produced using Monte Carlo experimentation. It is possible to measure the level of this variation using Kolmogorov and other statistics when the Monte Carlo experimentation involves completely random sampling. However, at present, this is not possible when other sampling schemes are involved. This means that the uncertainty inherent in Monte Carlo estimates cannot be properly measured when partial sampling schemes, such as Latin hypercube sampling, are applied.

The adjoint method can be used to characterise model output distributions in terms of approximate values of their moments, but not to calculate the distributions explicitly. The limitations and applicabilities of these approximate formulae deserve to be studied more rigorously because the adjoint method makes it possible to explore the behaviours of mathematical models over their entire parameter spaces.

These are at least two responses that we can make to this (apparent) limitation of the adjoint method. We can attempt to extend the method to make determinations of distribution functions and of their relatives possible. We can also look for ways to phrase regulations in terms of moments, or some other distribution characteristics, rather than the functions themselves.

It is possible to determine functions which are "envelopes" of model output distribution functions, using the distribution's moments. A simple example of this approach involving Chebchev's Inequality is included in this report for illustrative purposes. Other methods can be developed using sharper inequalities and other refinements.

The mathematical approximation errors present in adjoint method estimates (including infinite series truncation error, etc) are, in a way, analogous to the statistical sampling errors present in estimates obtained using Monte Carlo techniques. These errors are not reflections of the inherent uncertainties of mathematical modelling. Rather they arise because we are unable to calculate model output distributions exactly, and must therefore resort to approximations.

7.0 AREAS FOR FURTHER INVESTIGATION

- It is of the utmost importance that regulators be in possession of criteria and guidelines for determining whether proposed models have been properly designed, verified and qualified.

- The literature covering applications of the adjoint method to probabilistic treatments of model reliability should be reviewed. Limits of applicability, and pitfalls in applying the method, should be some of the results of such a study.

- All summaries of society's knowledge of parameter values, scenario probabilities, etc depend on expert opinion. Means for objectively elliciting, summarizing, reconciling and compiling expert opinion, as they apply to uranium mine tailings should be investigated. Regulators should be in possession of objective criteria for verifying the quality of this information.

- Mathematical means for assessing and characterizing scenario uncertainty and its consequences should be further investigated.

- As we have seen the appropriateness of any statistical or analytical method for a particular area of application depends on what estimates or inferences are required. What estimates or inferences are required in turn depends on the style of rules set forth in regulation and legislation. Thus, further inquiry should be made into the availability of mathematical methods for meeting the various kinds of information requirements which arise as consequences of the many styles of regulation, management, design and performance monitoring.

Appendix 1

APL programs to display empirical cumulative distribution functions and empirical frequency polygons, using a Commodore SuperPET SP9000 microcomputer to drive a Tektronix 4662 flatbed plotter:

Given a sample, and certain other information, the APL function PDF_DRAW listed below will provide a plot of a simple empirical probability density function. In addition to the sample vector, the program requires two other pieces of information: a "step size" which is the horizontal distance between successive points that are plotted, and a "smoothness factor" which controls the level of smoothing done in providing the estimator. If the smoothness factor is small in comparison to the sample range, then the estimator includes more "detail" (some of which, no doubt, will be noise). If the smoothness factor is a significant fraction of the sample range, then the estimate produced will be relatively "smooth".

The estimate shown as Figure A.1 was produced using:

```
.1  1  PDF_DRAW sample
```

where "sample" is the series of numbers presented in Section 3.1. (In fact, Figure A.1 presents an estimate of the probability density function shown as Figure 3.1.) The values .1 and 1 were chosen during a short series of trials - there is always some arbitrariness in applying statistical method, and this fact is especially obvious in estimation of density functions.

It must be emphasized that this program embodies one of the crudest methods for estimating probability density functions. The method resembles a Rosenblatt (56) kernel estimator. In practice the choice of an appropriate density function estimator must be based on some knowledge of the (unknown) density function, and, for reasons including this one, determination of suitable estimation methods for use in this context would make an interesting study in itself. For an introduction to probability density function estimation, see Wegman (72a,b).

Probability density function estimate

Figure A .1

```
[0]HS PDF_DRAW SAMPLE;X_S;Y_S;X;Y;H;DEL
[1]SAMPLE←SAMPLE[⍋SAMPLE]
[2]DEL←HS[1]
[3]H←HS[2]
[4]X_S←(SAMPLE[1]-H),(SAMPLE[⍴SAMPLE]+H),20 95
[5]Y_S←0,((⍴SAMPLE)÷H),25 65
[6]MOVE (X_S[1] SCALE X_S),0 SCALE Y_S
[7]DRAW (X_S[2] SCALE X_S),0 SCALE Y_S
[8]MOVE (X_S[1] SCALE X_S),0 SCALE Y_S
[9]X←SAMPLE[1]-H+DEL
[10]→E_L_1
[11]L_1: DRAW (X SCALE X_S),((+/(SAMPLE∘.(X+H)∧SAMPLE≥X)÷H) SCALE Y_S
[12]     E_L_1: →L_1×⍳(H+¯1↑SAMPLE)≥X←X+DEL
[13](↑X_S[1]) CENTRE_STRING 0,(X_S[1] SCALE X_S),23 1 1
[14](↑X_S[2]) CENTRE_STRING 0,(X_S[2] SCALE X_S),23 1 1
[15]'⍝PROBABILITY DENSITY FUNCTION ESTIMATE' CENTRE_STRING 0 55 10 1 1
[16]'_IGURE 5.4' CENTRE_STRING 0 55 5 1 1


[0]MOVE COORDS
[1]ONE_PLOT_XACTION UP_CASE 'M',(↑COORDS[1]),',',(↑COORDS[2])


[0]DRAW COORDS
[1]ONE_PLOT_XACTION UP_CASE 'D',(↑COORDS[1]),',',(↑COORDS[2])


[0]Z←SCALE ARGS
[1]Z←ARGS[3]+((X-ARGS[1])×ARGS[4]-ARGS[3])÷ARGS[2]-ARGS[1]
```

```
[0]Z←UP_CASE APL;BASICCHARS;APLCHARS
[1]BASICCHARS←',. +-1234567890QWERTYUIOPASDFGHJKLZXCVBNM'
[2]APLCHARS←',. -+1234567890?⍵∈⍴~↑↓⍳0×⌈⌊_∇∆∘'⍞⊂⊃∩∪⊥⊤|'
[3]→CANTDO×⍳1-⌊/APL⍷BASICCHARS
[4]Z←APLCHARS[BASICCHARS⍳APL]
[5]→0
[6]CANTDO: Z←''



[0]STRING CENTRE_STRING ARG;L;H;T;Y
[1]CHAR_PARMS (1.5 1.636 1)×ARG[4 5 1]
[2]H←ARG[5]
[3]T←(○ARG[1])÷180
[5]Y←ARG[3]
[4]Y←ARG[3]
[5]MOVE (ARG[2]-0.5×(L×2○T)-H×1○T),Y-0.5×((L←0.5×ARG[4]×¯1+3×⍴,STRING)×1○T)+H×2○
[6]PRINT STRING



[0]PRINT MSG
[1]ONE_PLOT_XACTION (UP_CASE 'P'),MSG


[0]ONE_PLOT_XACTION XACTION;I
[1]I←0
[2]ME: }
[2]ME: →ME×⍳(I←I+1)⍷⎕NUMS
[3]⎕←'IEEE9' ⎕CREATE I
[4]⎕←(⎕XR XACTION,⎕TC[7]) ⎕PUT I
[5]⎕UNTIE I
```

The APL function EDF_DRAW provides plots of the empirical distribution function for Latin hypercube sampling which was presented in Section 3.1. EDF_DRAW is listed below. Since it uses many of the same auxiliary routines for plotting as does PDF DRAW, these routines are not listed again with EDF_DRAW.

The plot which appears as Figure A.2 was produced using:

```
2 EDF_DRAW sample
```

where "sample" is the series of numbers presented in Section 3.1, and 2 is the notional number of parameters in the mathematical model, for this example only, which generated the sample. EDF_DRAW calls another APL function, PNI, which, given an argument I, returns I/N times the probability in the interval from which the Ith smallest observation in the sample was selected during Latin hypercube sampling.

Note that the graph in Figure A.2 achieves negative values. Although the estimator that EDF_DRAW embodies is unbiased, it does not exhibit all of the properties of the population exceedance function. That is, its range is not restricted to the interval $0,1$.

Appendix 2: Notes Concerning the Beta Distribution

If X is a beta random variable distributed over the interval (0,1) then the density function associated with X is given by

$$f(x;a,b) = x^{a-1}(1-x)^{b-1}/B(a,b) \quad 0 \leq x \leq 1;\ a,b > 0$$

where B(.,.) is the complete beta function

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

a and b are the density's parameters; B(a,b) is a normalizing constant.

Some of the simpler properties of the beta density are:

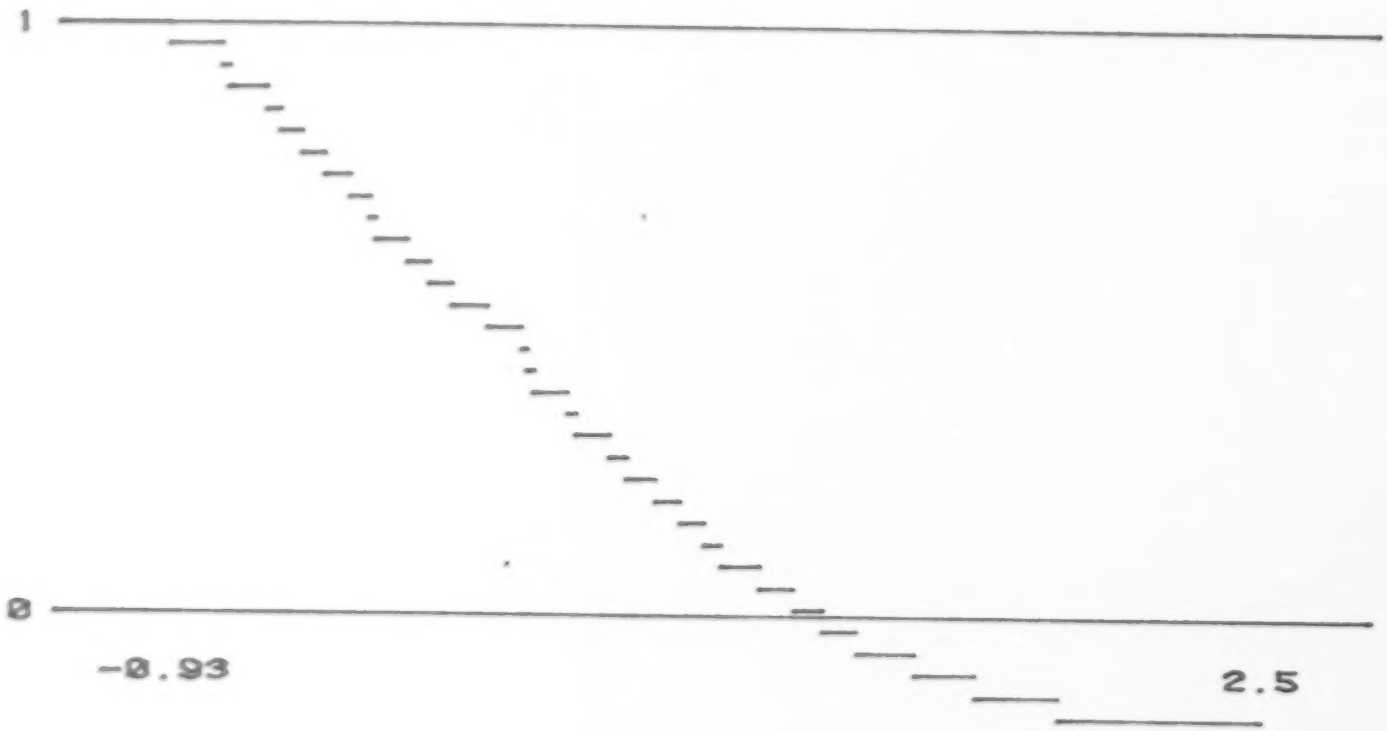
- symmetry in a and b: f(x;a,b) = f(1-x;b,a) ("if you've identified parameters that give a beta density that is skewed to the right, but you require one that is the mirror image of it, then just reverse the roles of the parameters.")

- (as a consequence of the previous property) the density is symmetrical about x=1/2 when a=b

- the density is U-shaped when both parameters are less than 1; otherwise it is unimodal, except when a=b=1, in which case it corresponds to the uniform density

- as the ratio a/b approaches zero the density becomes increasingly skewed to the left; conversely, as the ratio b/a approaches zero the density becomes increasingly skewed to the right.

A beta random variable distributed over the interval (0,1) can be called a standard beta variable. Non-standard beta variables (beta variables distributed over some finite interval other than (0,1)) can be constructed from standard beta variables by linear transformation.



Latin Hypercube

Empirical Survival Function

Figure A .2

```
[0]K EDF_DRAW SAMPLE;X_S;Y_S;X;Y;T;I;N;SPNI
[1]SAMPLE←SAMPLE[⍋SAMPLE]
[2]X_S←(SAMPLE[1]-T),(SAMPLE[⍳N]+T←0.1×SAMPLE[N←⍴SAMPLE]-SAMPLE[1]),20 95
[3]Y_S←¯0.2 1 25 65
[4]MOVE (X_S[1] SCALE X_S),0 SCALE Y_S
[5]DRAW (X_S[2] SCALE X_S),0 SCALE Y_S
[6]MOVE (X_S[1] SCALE X_S),1 SCALE Y_S
[7]DRAW (X_S[2] SCALE X_S),1 SCALE Y_S
[8]'1' CENTRE_STRING 0 18,(1 SCALE Y_S),1 1
[9]'0' CENTRE_STRING 0,18,(0 SCALE Y_S), 1 1
[10](⍕TOF ⍎SAMPLE[1]) CENTRE_STRING 0,(SAMPLE[1] SCALE X_S),(¯0.1 SCALE Y_S),1 1
[11](⍕TOF ⍎SAMPLE[N]) CENTRE_STRING 0,(SAMPLE[N] SCALE X_S),(¯0.1 SCALE Y_S),1 1
[12]X←X_S[1]
[13]SPNI←I←0
[14]MOVE (X_S[1] SCALE X_S),(1-(N×K-1)×SPNI) SCALE Y_S
[15]→E_L_1
[16]L_1: DRAW (SAMPLE[I] SCALE X_S),(1-(N×K-1)×SPNI) SCALE Y_S
[17]     SPNI←SPNI+PNI I
[18]     MOVE (SAMPLE[I] SCALE X_S),(1-(N×K-1)×SPNI) SCALE Y_S
[19]     E_L_1: →L_1×⍳(N≥I←I+1)
[20]'DATIN ATPERCUBE' CENTRE_STRING 0 55 15 1 1
[21]'EMPIRICAL FURVIVAL _UNCTION' CENTRE_STRING 0 55 10 1 1
```

```
[0]Z←PNI I
[1]Z←1÷30×30
```

Appendix 3: Bounds on Probabilities via Chebychev's Inequality

We use Chebychev's Inequality to provide a simple illustration of the fact that values for distribution moments - in this case the mean and variance - can be used to give bounds on exeedance probabilities.

In this report we are primarily concerned with the probability that a given measure of consequence - dose (say) - will exceed a given arbitrary limit. That is, if D is dose, and if d is the arbitrary limit, then we require $\Pr\{D \geq d\}$. Let us suppose that we know the mean, μ, and variance, σ^2, for the distribution on dose but that we do not have a complete description of the distribution itself.

Chebychev's Inequality for this situation is

$$\Pr\{D \geq \mu + \delta\sigma\} < 1/\delta^2,$$

Where δ is an arbitrary constant.

If we set $d = \mu + \delta\sigma$ then $\delta = (d - \mu)/\sigma$. So $\Pr\{D \geq d\} < \sigma^2/(d-\mu)^2$.

Thus, Chebychev's Inequality gives an upper bound on $\Pr\{D \geq d\}$ in terms of the mean and variance of the distribution on D. Since $\Pr\{D \leq d\}$ is the value of the exceedance function on D at d this is also an upper bound on the value of the exceedance function.

In practice one would not use the Chebychev inequality, in the simple form presented here, to turn information about means and variances into information about exceedance probabilities. However, this should serve as an example.

REFERENCES

IEC BEAK (82). An Approach to Long Term Aquatic Modelling of a Generic Uranium Tailing Site, A Report for AECB, IEC BEAK Consultants Ltd.

M.S. Chu and R.M. Cranwell (81). "Structure and overview of the Sandia waste isolation risk assessment methodology", in Proc of the Symposium on Uncertainties Assoc. with the Regulation of the Geologic Disposal of High-Level Radioactive Waste, NUREG/CP-0022-Conf-810372.

J. Coady (82). Telephone and telex communications.

R.M. Cranwell and J.C. Helton (81). "Uncertainty analysis for geologic disposal of radioactive waste", in Proc of the Symposium on Uncertainties Assoc. with the Regulation of the Geologic Disposal of High-level Radioactive Waste, NUREG/CP-0022-Conf-810372.

W. Dormuth and G.R. Sherman (81). SYVAC - A Computer program for Assessment of Nuclear Fuel Waste Management Systems, Incorporating Parameter Variability, AECL-6814.

B. de Finetti (70). "Logical foundations and measurement of subjective probability", Acta Psychologica 34, pp 129-45.

R.H. Gardner, D.D. Huff, R.V. O'Neill, J.B. Rankin, J. Carney and J.J. Jones (80). "Application of error analysis to a marsh hydrology model", Water Resour. Res. 16, pp. 659-64.

R.H. Gardner, R.V. O'Neill and F.O. Hoffman (81). "Assessing model uncertainties" in Proc. of the Symposium on Uncertainties Assoc. with the Regulation of the Geologic Disposal of High-level Radioactive Waste, NUREG/CP-0022-Conf-810372.

M.S. Giuffre and J.Y. Nalbandian (81). "Methods for determining system uncertainty", in Proc of the Symposium on Uncertainties Assoc with the Regulation of the Geologic Disposal of High-level Radioactive Waste, NUREG/CP-0022-Conf-810372.

N.A.J. Hastings and J.B. Peacock (75). Statistical Distributions, Halsted Press.

C.R. Heathcote (71). Probability: Elements of the Mathematical Theory. Geo. Allen and Unwin. London.

R.M. Hogarth (77). "Cognitive processes and the assessment of subjective probability distributions", in New Developments in the Applications of Bayesian Methods. No. Holland Publishing. also Journal of the American Statistical Association 70, pp. 271-89.

L.W. Huson (82). "A graphical aid to multivariate sensitivity analysis", Ecological Modeling 16, pp 91-8.

R.L. Iman (81). "Statistical methods for including uncertainties associated with the geologic isolation of radioactive waste which allow for comparison with licensing criteria", in Proc. of the Symposium on" Uncertainties Assoc. with the Regulation of the Geologic Disposal of High-level Radioactive Waste, NUREG/CP-0022-Conf-810372.

R.L. Iman and W.J. Conover (80). "Small sample sensitivity analysis technicques for computer models, with an application to risk assessment", Commun Statist. Theor. Meth. A9(17). pp. 1749-842.

R.L. Iman , J.C. Helton and J.E. Campbell (79). "Risk methodology for geologic disposal of radioactive waste: sensitivity analysis techniques", "U.S. Nuclear Regulatory Commission Report, NUREG/CR-0394.

E.T. Jaynes (59) "Probability theory in science and engineering", Colloquium Lectures in Pure and Applied Science, No. 4, Socony-Mobil Oil Company.

B. Kadane, J.M. Dickey, R.L. Winkler, W.S. Smith and S.C. Peters (80). "Interactive ellicitation of opinion for a normal linear model". Journal of the American Statistical Association 75, pp. 845-54.

M.G. Kendall and A. Stuart (73). The Advanced Theory of Statistics. Volume 2: Inference and Relationship, 3rd ed. Griffin and Co.

U.W. Kulisch and W.L. Miranker (81). Computer Arithmetic in Theory and Practice. Academic Press.

M.D. McKay, W.J. Conover and R.J. Beckman (79). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code", Technometrics 21, pp 239-45.

R.V. O'Neill, R.H. Gardner and J.H. Carney (82). "Parameter constraints in a stream ecosystem model: incorporation of a priori information in Monte Carlo error analysis", Ecological Modeling 16, pp 51-65."

L.D. Phillips (70) "The 'true' probability problem", Acta Psychologica 34, pp 254-64.

M. Rosenblatt (56). "Remarks on some nonparametric estimates of a density function", Ann. Math. Stat. 27, pp. 832-7.

B. Ross, C.M. Koplik (78). "A statistical approach to modelling transport of pollutants in groundwater", Mathematical Geology 10, pp 657-72."

R.E. Schaeffer and K. Borcherding (73). "The assessment of subjective probability distributions: a training experiment", Acta Psychologica 37, pp 117-29.

S.A. Schmitt (69). Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics, Addison-Wesley.

C.A. Stael von Holstein (71). "Two techniques for assessment of subjective probability distributions - an experimental study", Acta Psychologica 35, pp 478-94.

(70). "Measurement of subjective probability", Acta Psychologica 34, pp 146-59.

R.E. Thomas (Mar 82). "Uncertainty analysis", Technical Report ONWI-380: Battelle Columbus Laboratories, Columbus, Ohio.

M. Tribus (62). "The use of the maximum entropy estimate in the estimation of reliability", in R.E. Machol and R. Gray (eds.), Recent Developments in Information and Decision Processes, Macmillan, New York.

R.L. Winkler (67). "The assessment of prior distributions in Bayesian statistics". Journal of the American Statistical Association 62, pp 776-800.

E.J. Wegman (72a) "Nonparametric probability density estimation, I. A summary of available methods," Technometrics 14, pp 533-46.

E.J. Wegman (72b) "Nonparametric probability density estimation,II. A comparison of density estimation methods," Journal of Statistical Computation and Simulation 1, pp 225-45.

(67) "The quantification of judgement: some methodological suggestions", Journal of the American Statistical Association 62, pp 1105-20.

(67) "The quantification of judgement: some experimental results" Proceedings of the American Statistical Association, pp 306-95.

(80) "Prior information, predictive distributions, and Bayesian model-building", In Bayesian Analysis in Econometrics and Statistics: "Essays in Honor of Harold Jeffreys, North Holland Publishing, Amsterdam.

R.L. Winkler and S. Murphy (68) "Good probability assessors", Journal of Applied Meteorology 7, pp 751-8.